CONFIDENTIAL UNTIL PUBLISHED
Evidence Review Group's Report
Tofacitinib for treating active psoriatic arthritis following disease modifying anti-rheumatic drugs

Produced by CRD and CHE Technology Assessment Group, University of York, Heslington, York YO10 5DD

Authors Ruth Walker, Research Fellow CRD

Aimee Fox, Research Fellow CHE

Marta Soares, Senior Research Fellow

Nerys Woolacott, Reader Emerita, CRD

Nick Meader, Research Fellow, CRD

Melissa Harden, Information Specialist, CRD

Stephen Palmer, Professor of Health Economics, CHE

Laura Bojke, Senior Research Fellow, CHE

Correspondence to Laura Bojke, Senior Research Fellow, CHE, University of York YO10 5DD. Tel: (01904) 321416. Email:laura.bojke@york.ac.uk

Date completed Date completed (14/06/2018)

Source of funding

This report was commissioned by the NIHR HTA Programme as project number 17/109/05.

Declared competing interests of the authors

None.

Acknowledgements

We would like to thank Dr Pauline Ho (Consultant Rheumatologist) of the Kellegren Centre for Rheumatology, Central Manchester and Manchester Children's University Hospital Trust, for clinical advice throughout the project.

Rider on responsibility for report

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

This report should be referenced as follows:

Walker R, Fox A, Soares M, Woolacott N, Meader N, Harden M, Palmer S, Bojke L. Tofacitinib for treating active psoriatic arthritis: A Single Technology Appraisal. CRD and CHE Technology Assessment Group, 2018.

Contributions of authors

Ruth Walker, Nerys Woolacott, and Nick Meader wrote the clinical effectiveness section. Aimee Fox, Marta Soares and Laura Bojke wrote the cost-effectiveness sections of the report and conducted the economic analyses. Melissa Harden wrote the sections on the search strategies. Stephen Palmer provided advice and commented on drafts of the report. Laura Bojke took overall responsibility for the cost-effectiveness sections. Nerys Woolacott took overall responsibility for the clinical effectiveness sections of the report.

Note on the text

All commercial-in-confidence (CIC) data have been highlighted in blue and underlined, all adademic-in-confidence (AIC) data are highlighted in yellow and underlined

Copyright statement

Table of Contents

Table of Tables

Table of Figures

List of abbreviations

ACR	American College of Rheumatology
ACR20/50/70	20%/50%/70% improvement in the ACR response criteria
AE	Adverse event
bDMARD	Biological disease-modifying anti-rheumatic drug
BSC	Best supportive care
CEA	Cost-effectiveness analysis
CHE	Centre for Health Economics
CHMP	Committee for Medicinal Products for Human Use
CrI	Credible interval
CSR	Clinical study report
csDMARD	Conventional synthetic disease-modifying anti-rheumatic drug
DMARD	Disease-modifying anti-rheumatic drug
EMA	European Medicines Agency
EQ-5D	5-dimension European Quality of Life questionnaire
ERG	Evidence Review Group
HAQ-DI	Health Assessment Questionnaire-Disability Index
HRQL	Health-related quality of life
ICER	Incremental cost-effectiveness ratio
JAK	Janus Kinase
LOCF	Last observation carried forward
MCID	Minimum clinically important difference
MS	Manufacturer’s submission
MXT	Methotrexate
NIHR	National Institute for Health Research
NMA	Network meta analysis

NR	Not reported
NRI	Non-responder imputation
PALACE	Psoriatic Arthritis Long-term Assessment of Clinical Efficacy
PASI	Psoriasis Area and Severity Index
PASI-50/75/90	50%/75%/90% or greater improvement in PASI score
PsA	Psoriatic arthritis
PsARC	Psoriatic Arthritis Response Criteria
QALY	Quality-adjusted life year
RCT	Randomised controlled trial
SF36	36-item Short-Form Health Survey
SJC	Swollen joint count
STA	Single Technology Appraisal
TNF	Tumour necrosis factor
tsDMARD	disease-modifying anti-rheumatic drugs targeting a particular molecular structure
WTP	Willingness to pay

1 Summary

1.1 Critique of the decision problem in the company's submission

Tofacitinib is an oral, small molecule, targeted Janus Kinase (JAK) inhibitor. A positive opinion from the Committee for Medicinal Products for Human Use (CHMP) was adopted in April 2018 for the use of tofacitinib 5mg BD, twice daily,

''in combination with methotrexate (MTX) for the treatment of active psoriatic arthritis (PsA) in adult patients who have had an inadequate response or who have been intolerant to a prior disease-modifying antirheumatic drug (DMARD) therapy''.

The NICE scope differed from the licence in that tofacitinib could be used alone or in combination with non-biological DMARD. The CS assessed tofacitinib in combination with any csDMARD and did not restrict to the use of MTX.

The CS addressed three sub-populations, those who had not adequately responded to at least two non-biologic DMARDS, those who had not adequately responded to non-biologic DMARDS and one or more tumour necrosis factor inhibitors (TNFis), and those for whom TNFis are contradicted or not tolerated. The CS did not include a fourth sub-population that had been included in the NICE scope (those who had failed one non-biological DMARD) as there were insufficient data.

The comparators addressed in the company's decision problem matched those in the NICE scope for (1) those who had not adequately responded to at least two non-biologic DMARDS and (2) those for whom TNFis are contradicted or not tolerated. For the subpopulation, those who had not adequately responded to csDMARDS, certolizumab pegol was not addressed. The ERG agreed with the exclusion of this comparator as the RAPID PsA trial did not include all TNFi experienced patients, but only those who had initially responded to a TNFi and then lost their response [1].

1.2 Summary of clinical effectiveness evidence submitted by the company

The clinical effectiveness evidence for the use of tofacitinib in active PsA consisted of two placebo-controlled RCT's; one for TNFi naïve (OPAL Broaden) and one for TNFi experienced patients (OPAL Beyond). Patients from these trials who received tocacitinib 10mg BD doses did not contribute to the clinical effectiveness evidence submitted by the company, as the use of tofacitinib is licenced for dose 5mg BD. Supporting evidence from a non-RCT open-label follow-up study of tofacitinib, OPAL Balance, was also presented.

OPAL Broaden and OPAL Beyond were well conducted Phase III randomised, multicentre trials. OPAL Broaden also included a comparison with adalimumab and after 3-months, patients receiving placebo were followed up on tofactinib or adalimumab to 12 months. OPAL Beyond did not include a comparison with adalimumab and after the 3 months, patients receiving placebo were followed up on tofactinib to 6 months.

Baseline characteristics were similar across both trials. The primary efficacy outcomes were ARC20 response rate at 3 months and ΔHAQ-DI at 3 months. Modified PsARC response and PASI-75 response were also included as outcomes. Radiographic assessment of joint damage was also assessed at 12 months within OPAL Broaden.

TNFi naive population

For TNFi naïve patients, OPAL Broaden demonstrated that tofacitinib 5mg BD (N= 107) was statistically significantly more effective than placebo (N=105) for the key efficacy outcomes; ARC 20/50/70, PASI70 response rate and mean ΔHAQ-DI at 3 months, but not PSARC response rate. Comparisons of tofacitinib with adalimumab show that numerically for most key efficacy outcomes adalimumab was very slightly better than tofacitinib, however the trial was not powered to test for a statistically significant difference or non-inferiority. For radiographic assessment of joint damage the proportion of progressors (change in mTSS of >0.5) was low in both treatment arms but the upper confidence interval in the population adjusted analyses (to be comparable with the ADEPT trial for adalimumab) crossed the non-inferiority margin +ndicating it was inconclusive whether tofacitinib 5mg was non-inferior to adalimumab. The ERG agreed with the FDA conclusion that there is insufficient evidence to support the assumption that tofacitinib is associated with halting radiographic progression.

Network meta-analyses across outcomes (e.g. PsARC, ACR, PASI, and HAQ changes conditional on PsARC response) found that golimumab, infliximab, and etanercept were generally the most effective treatments; followed by certolizumab, secukinumab 150, adalimumab, and secukinumab 300. Apremilast, ustekinumab and tofacitinib 5mg were consistently ranked among the lowest in effectiveness. The company found that the placebo arm of OPAL Broaden fitted poorly in their NMA models, and attributed this to the high placebo response observed in their trial. They therefore presented alternative analyses including one where the placebo arm of OPAL Broaden was excluded which also resulted in increased effectiveness estimates for tofacitinib 5mg.

TNFi experienced population

For TNFi experienced patients, OPAL Beyond demonstrated that tofacitinib 5mg BD (N= 131) was statistically significantly more effective than placebo (N=131) for the key efficacy outcomes

outcomes; ACR 20/50, PsARC response rate and mean ΔHAQ-DI at 3 months but not ACR 70 or PASI 75.

Network meta-analyses for PsARC and HAQ changes conditional on PsARC response only included ustekinumab and tofacitinib and were found to be of similar effectiveness. Tofacitinib was associated with only slightly higher HAQ changes than placebo. More treatments were included for PASI response, the results of which found that tofacitinib 5mg was among the least effective in the network meta-analysis: ustekinumab, secukinumab, ixekizumab were ranked higher and only abatacept was ranked less effective.

Adverse effects

The adverse events profile of tofacitnib in PsA patients appears similar to, and no worse than that of adalimumab. The tolerability of tofacitinib is reflected in the low rate of withdrawals due to AEs. An increased risk of herpes zoster appears to be a specific AE of tofacitinib.

1.3 Summary of the ERG's critique of clinical effectiveness evidence submitted

The evidence for the clinical effectiveness of tofacitinib is based on good quality randomised trials and the results are likely to be reliable.

The ERG identified limitations in the generalisability of the RCT evidence to clinical practice. These were owing to a significant proportion of patients in each RCT (18% and 24%) treated in combination with sulfasalazine and leflunomide, when the marketing authorisation is for tofacitinib in combination with methotrexate (MTX) only. Furthermore, in both OPAL Broaden and OPAL Beyond the placebo-controlled phase was limited to 3 months: treatment with tofacitinib in clinical practice is long-term. Additional issues relating to generalisability included:

(1) The use of adalimumab in OPAL Broaden in combination with a csDMARD not being reflective of adalimumab in clinical practice or in other trials.
(2) The number of previous TNFis (and the specific previous TNFis) in OPAL Beyond not being reflective of the patient population in which tofacitnib will be used in current practice.
(3) [redacted], whereas the licenced dose for tofacitinib in 5mg BD.

The ERG identified errors in the implementation of the company's placebo-adjusted NMAs. Models corrected by the ERG found a more meaningful interaction between baseline risk and treatment effect than the company analyses.

1.4 Summary of cost effectiveness submitted evidence by the company

The CS submitted a decision model, which allows the comparison of multiple treatment sequences to evaluate the cost-effectiveness of tofacitinib.

The population included people whose disease has not responded adequately to two non-biological DMARDs, people whose disease has not responded adequately to non-biological DMARDs and one or more TNFis and people in whom TNFi are contraindicated or not tolerated.

For all outcomes (PsARC response, PASI response, and HAQ-DI change conditional on PsARC response), response rates for tofacitinib 5 mg BD and its comparators were taken from the network meta-analyses (NMAs), where available. Patients in the model were assumed to continue with therapy after 12 weeks if they achieve PsARC response and HAQ and PASI were assumed constant (no disease progression) for those that have a PsARC response. Withdrawal from therapy at any point after primary response was assumed to be the same for tofacitinib and all comparators. HRQoL and costs were a function of HAQ and PASI score, in addition to the costs of medication, administration and monitoring. The acquisition costs of treatments were estimated from the British National Formulary.

The original model was revised following requests for clarifications from the ERG. In their response to clarification, the company identified a data entry error for Models K1 and K2 in the bDMARD-naïve NMA. They rectified the error in the revised and corrected version. The revised version also included an updated PAS price for tofacitinib that had been approved since the original CS.

The revised results from the CS suggest that the tofacitinib 5 mg BD sequence may be a cost-effective option (at conventional willingness to pay thresholds) vs BSC for each sub-population. In each of the three sub-populations assessed, the deterministic ICER for tofacitinib 5 mg BD vs BSC was below £20,000 per QALY.

1.5 Summary of the ERG's critique of cost effectiveness evidence submitted

The ERG had a number of concerns regarding assumptions and data used in the CS and economic model. In particular, the assumption that tofacitinib halts HAQ-DI progression while patients remain on treatment. The ERG is cautious of this assumption given that no long-term clinical evidence is available to support this, such as data assessing radiographic disease progression.

The ERG also had concerns about assumptions made regarding effect degradation for subsequent lines of therapy. Subsequent treatments are assumed to be as efficacious as first line, i.e. no effect

degradation is assumed. Due to the lack of flexibility in the company model, the ERG is unable to explore the sensitivity of the cost-effectiveness results to this assumption.

The ERG found errors in the NMA placebo-response adjusted models and concluded that these were incorrectly implemented. The ERG corrected the company base-case model and revisited model selection to select the ERG's preferred base-case.

1.6 ERG commentary on the robustness of evidence submitted by the company

1.6.1 Strengths

Clinical Effectiveness

The CS included a systematic review of the evidence for tofacitinib and all relevant comparators and also built on a previous NICE MTA (TA445). The evidence for clincial effectiveness was derived from two well conducted RCTs, one each for TNFi naïve and –experienced patients. The trial in TNFi- naïve patients also included a comparison with adalimumab, which was very informative: in studies of bDMARDS for PsA direct comparisons with active treatments are infrequently made. To compare tofacitinib with the long list of relevant comparators appropriate NMA were conducted.

Cost effectiveness

A de novo model based on previous NICE technology appraisals was developed. This uses a model structure similar to that developed for TA445 and utilises much of the same data and assumptions. The CS presented a de novo NMA which incorporates all relevant clinical evidence for all comparators.

1.7 Weaknesses and areas of uncertainty

Clinical Effectiveness

As outlined in Section 1.3 above, the included trials had some limitations in their generalisability to clinical practice. Longer term data are required to confirm the efficacy of tofacitinib, particularly for the outcome of progression of joint disease. The trial was not powered to test whether tofacitinib was non-inferior to adalimumab and was therefore inconclusive.

Cost effectiveness

There are a number of parameter uncertainties within the company's model. The most critical of these is the assumption of zero HAQ-DI progression for PsARC responders to tofacitinib remaining on treatment, without radiographic or randomised trial data sufficient to support this assumption. The ERG also had concerns on assumptions regarding: no effect degradation, the psoriasis sub-groups and the impact of other approved PAS prices.

1.8 Summary of exploratory and sensitivity analyses undertaken by the ERG

The ERG conducted a range of exploratory analyses to assess the uncertainties raised in the review and critique of the manufacturer's clinical and cost-effectiveness evidence. The ERG's exploratory analyses focused on, severity of psoriasis, tofacitinib progression rates and drug costs for comparator drugs that are approved but not available publicly. The additional analyses undertaken by the ERG suggested that whilst the ICERs for all subpopulations changed in each of the scenarios, they remained within the acceptable willingness to pay threshold, compared to BSC (typically below £20,000 per QALY). The fully incremental ICERs for tofacitinib and etanercept are also within conventional willingness to pay thresholds.

2 Background

2.1 Critique of company's description of underlying health problem

The description of the underlying health problem in the company's submission was appropriate and relevant to the decision problem under consideration. Psoriatic arthritis (PsA) is an inflammatory condition with onset usually occurring between 30 and 50 years of age. Clinical manifestations are heterogeneous and may include both articular (joint) and non-articular disease features. The CS states patients have an onset of psoriasis occurring 7 to15 years prior their PsA diagnosis [2]. PsA is a chronic, progressive condition leading to irreversible joint damage and is additionally associated with a range of comorbidities including hypertension, hyperlipidaemia, depression, fibromyalgia and type II diabetes [3]. The five health domains of pain (in the joints and spine), skin problems (including itching), fatigue (both physical and mental), ability to pursue work and leisure activities, and functional capacity are identified as the most important from the patients' perspective [4].

2.2 Critique of company's overview of current service provision

The manufacturers' overview of current service provision is broadly appropriate and relevant to the decision problem under consideration. NICE clinical guidance (NG65) is outlined in the CS and in full in CS Appendix L; in addition guidance from the European League Against Rheumatism (EULAR), the British Society for Rheumatology (BSR) and the Group for Research and Assessment for Psoriasis and Psoriatic Arthritis (GRAPPA) is also detailed in the CS. Clinical guidelines for PsA emphasise the control of symptoms, prevention of structural damage, and normalisation of functional and social participation and propose disease remission or low/ minimal disease activity as the therapeutic treatment goal.

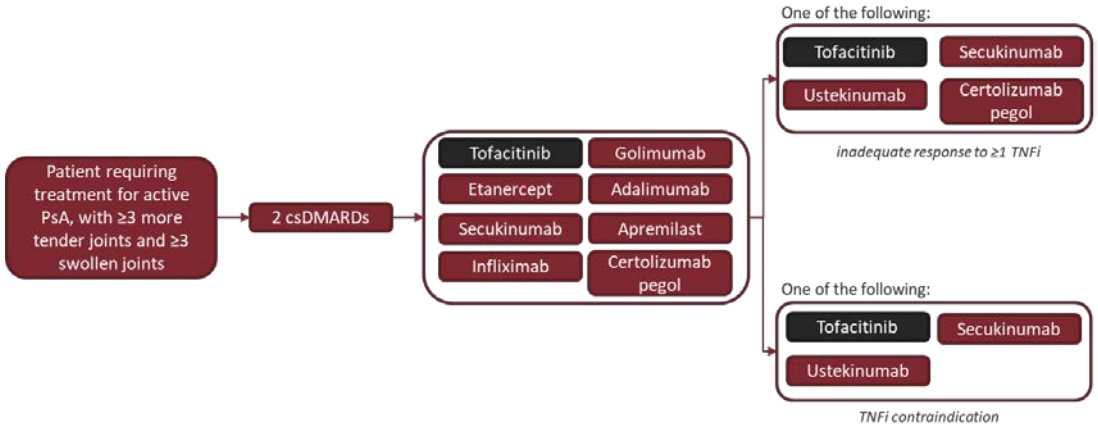
The CS states the proposed positioning of tofacitinib (Figure 1) is after conventional synthetic disease-modifying anti-rheumatic drugs (csDMARDs) as an alternative to other currently recommended biologic disease-modifying anti-rheumatic drugs (bDMARD/tsDMARDs), after treatment failure or for those intolerance or contraindication to tumour necrosis factor alpha inhibitors (TNFi).

The rationale in the CS for the position of tofacitinib as an alternative to other currently recommended treatment options, for patients with active PsA who have had an inadequate response to previous treatments (csDMARDS and TNFis), was made on the basis of providing a treatment with the following characteristics:

- Oral route of administration
- A novel mechanism of action

- A proven efficacy profile across multiple PsA domains
- An acceptable safety profile

Figure 1 : Proposed Positioning of Tofacitinib in the Treatment Pathway (CS Figure 1)



The CS includes a section on problems associated with current use of bDMARDs in clinical practice (p23-24 CS). The problems highlighted are: patients' dissatisfaction, limitations associated with administration by injection or infusion, and sub-optimal treatment persistence associated with current therapies. The CS states there is a need for an (additional) oral treatment option for TNFi-naïve and TNFi-experienced patients.

The evidence in CS Section B.1.3.3 was largely taken from the Multinational Assessment of Psoriasis or Psoriatic Arthritis (MAPP) survey of patients. It is cited in the CS that bDMARDs were burdensome primarily due to the fear and anxiety associated with injections and the physical preparation for self-injection (26%), inconvenience (15%), adverse events (15%), pain/discomfort (7%), and a lack/loss of effectiveness (2%), and that 85% of patients report a need for better therapies for the treatment of PsA. The ERG notes that in the MAPP survey, only 21% of participants contributing to the 'treatment burden' outcome had PsA. Furthermore, the ERG notes that on the whole, the evidence from this survey does not fully support the suggestion that oral therapies are any better tolerated than other biologic therapies, or that injection site reaction, needle fatigue or injection anxiety played a major role in the discontinuation of treatments administered subcutaneously or via infusion. The ERG also notes that in the MAPP study, overall discontinuation rates were higher with

traditional oral therapies compared to biologic therapies (57% to 45%), with reasons for discontinuation between the two being similar: safety, tolerability or a lack or loss of effectiveness.[5] In a previous ERG report for the oral therapy apremlist for PsA, the ERG also noted that as the MAPP study was a based on a community cohort, rather than a pure hospital cohort where more severe disease is likely, the direct applicability of these findings is questionable: patients with milder symptoms are unlikely to tolerate the adverse effects of treatment as well as patients with more severe symptoms or disease. Furthermore, the number of UK patients in the survey was small (around 12%).

The CS also cites evidence from a U.S based survey of 468 patients. This choice-based conjoint survey determined patient preferences for treatment modalities for PsA and was mailed to 2,800 randomly selected patients enrolled in Humana Inc. Medicare and commercial plans (response rate 16.7%). Across both types of health plan, oral formulation was preferred relative to self-injection and intravenous routes of administration, and lower cost formulations were preferred. Results from this survey are available only in abstract form and average importance scores are presented, where the average score for 'route of administration' is highest for Medicare patients and average score for 'cost to you' is highest for commercial patients. The extent to which these findings are generalizable to UK patients is unknown.[6]

The clinical advisor to the ERG thought that oral treatment was not likely to be an important advantage from a patient's perspective. Whilst treatment requiring infusion such as infliximab, has the potential to be more burdensome to PsA patients, biological therapies requiring self-administered weekly or bi-weekly subcutaneous injection (etanercept administered once weekly and adalimumab administered once every two weeks),[7] may be less so. Furthermore, adherence and compliance with twice-daily tablets may well be poorer than to less frequent injections, and the clinical monitoring of adherence to tablets likely to more difficult than that of adherence to biologic therapies. Considering this, the need for an (additional) oral medication option for the treatment of PsA may not be as pressing as the CS suggests. In addition, the ERG notes that due to the requirment for tofacitinib to be given concomitantly with MTX (which many patients self-administer as a subcutaneous injection), treatment will not necessarily avoid an injection-based administration.

The CS states that among patients treated with TNFis, treatment persistence is low owing to a lack of response and, or tolerance to TNFis, implying the need for interventions with alternative mechanisms of action in TNFi-IR patients. The CS states that 30-50% of patients discontinue their index TNFi during the first treatment year. The CS cites evidence from a Danish cohort (2000-2009); stating 44% of patients discontinued their index TNFi therapy during the first year. The ERG notes that the cited figure of 44% of patients refers to those who discontinued TNFi therapy over the whole course of the study (median follow-up 2.9 years). One-year drug survival was in fact 70%, with two-year survival

57% [8]. The CS also states that the British Society of Rheumatology Biologics Register (BSRBR) indicates that only 59% of patients remain on their first TNFi for PsA after three years of treatment.[9]The ERG identified a recently published analysis of the UK based BSRBR data (625 PsA patients), which reported long-term persistence of etanercept, infliximab and adalimumab at 3, 5 and 8 years. Etanercept and adalimumab rather than infliximab were associated with better five-year persistence. At five years 46.7% were still on their initial TNFi treatment. Furthermore, at eight years, 33% remained on the first TNFi, 16% on the second and 12% on the third, and only 5% of patients were on a non-TNFi biologic and 10% not on a biologic treatment [10]. This suggests that within the UK, whilst patients may switch treatments, discontinuation from all biologic therapy is low at 8 years. In TNFi-IR patients, the extent to which issues with drug survival translate into the requirement for additional treatments options may be less than the CS suggests.

The CS also states that tofacitinib, as a small molecule JAK inhibitor would not be expected to induce any immunogenicity, as is associated with infliximab and adalimumab. Additional justification for this was provided in the company's response to points for clarification. This stated that the lack of association with immunogenicity was due to the lower molecular weight of tofacitinib compared to bDMARD's. The clinical advisor to the ERG advised that in clinical practice immunogenicity is not a significant issue.

Overall, the ERG acknowledges the novel mode of action of tofacitinib, but suggests that the company may have overstated the need for an oral treatment option for PsA. The efficacy relative to existing therapies is probably the key factor when deciding whether or not to use tofacitinib.

The clinical advisor to the ERG suggested that given there is limited knowledge of the use of ofacitinib in clinical practice, it would likely be reserved for an end of line treatment or possibly for specific individuals with certain clinical characteristics, for whom TNFis are contraindicated or not tolerated.

3 Critique of company's definition of decision problem

3.1 Population

The population stated in the CS was:

'Adults with active PsA whose disease has not responded adequately to previous DMARD therapy or for whom DMARDs are not tolerated or contradicted'.

This matches the NICE scope and accurately reflects the marketing authorisation.

3.2 Intervention

The intervention stated in the CS was:

'Tofacitinib (in combination with a csDMARD)'

This differs from the NICE scope that states 'tofacitinib (alone or in combination with an csDMARD)'. The marketing authorisation is for tofacitinib in combination with methotrexate (MTX) only. The clinical effectiveness of tofacitinib was informed by trials some including patients who were treated in combination with sulfasalazine and leflunomide. The licenced dose of tofacitinib is 5mg BD twice daily.

3.3 Comparators

The comparators stated in the CS are for three sub –populations (sub –populations 2, 3 and 4):

2 – For people whose disease has not responded adequately to at least 2 non-biological DMARDs: bDMARDs; apremilast; best supportive care.

3 – For people whose disease has not responded adequately to non-biological DMARDs and 1 or more TNFis: ustekinumab; secukinumab; best supportive care.

4 – For people in whom TNFis are contraindicated or not tolerated: ustekinumab; secukinumab; best supportive care.

This differs from the final scope issued by NICE that included comparators separately for one additional sub –population (sub –population 1):

1 – For people whose disease has not responded adequately to 1 non-biological DMARD

- Non-biological DMARDs

The CS states there was insufficient data to subdivide data from patients who had failed 1 non-biological DMARD and those who had failed 2 non-biological DMARDs (sub –population 1 and sub-population 2 in the NICE scope). Therefore, the company has not included this population in the submission. The ERG agrees with this and thinks it is reasonable.

Comparators for sub-populations 2, 3 and 4 in the decision problem addressed in the CS match those stated in the final NICE scope, expect for certolizumab pegol, which has been excluded from sub-population 3. The CS states this is because the data available from the RAPID PsA trial informs only a subset of patients in this sub-population. The ERG agrees with this: RAPID PsA did not include all

TNFi experienced patients, but only those who had initially responded to a TNFi and then lost their response.[1]

3.4 Outcomes

The outcome measures included in the decision problem addressed by the company were:

- Disease activity: ACR20, ACR50, ACR70, ACR response criteria components, PASI50/75/90, PsARC, MDA
- Functional capacity: HAQ-DI, HAQ-DI conditional on PsARC response status
- Disease progression: van der Heijde-mTSS
- Periarticular disease (for example, enthesitis, tendonitis, dactylitis): DSS, LEI, SPARCC
- Health-related quality of life: SF-36 (physical functioning component), FACIT-F (total score), DLQI, ISI
- Mortality
- Adverse effects of treatment
- EQ-5D: provided in the company's clarification response

These are consistent with those in the final scope issued by NICE.

3.5 Other relevant factors

No equity issues are anticipated should tofacitinib be recommended for used in England and Wales.

The patient access scheme (PAS) will provided a simple discount of [redacted] (discounted price of £[redacted] per 5mg 56-tablet pack) to the list price of tofacitinib, with the discount applied at the point of purchase or invoice.

4 Clinical Effectiveness

This section contains a critique of the methods of the review(s) of clinical effectiveness data, followed by a description and critique of the trials included in the review, including a summary of their quality and results and the results of any synthesis of studies.

4.1 Critique of the methods of review(s)

The CS included a systematic review across the intervention of interest (tofacitinib 5 mg) and all relevant comparators. The methods of the review are discussed in the sections below.

4.1.1 Searches

The search strategy used by the company to identify 1) relevant clinical data on the use of tofactinib for the treatment of PsA and 2) relevant clinical data regarding the clinical effectiveness of other existing treatments for PsA to be used in a network meta-analysis (NMA), were described in full detail in Appendix D.

The electronic databases MEDLINE, MEDLINE Daily, MEDLINE In Process, EMBASE and the Cochrane Library (including the Cochrane Database of Systematic Reviews (CDSR), the Database of Abstracts of Reviews of Effects (DARE), the Cochrane Central Register of Controlled Trials (CENTRAL), NHS Economic Evaluation Database (NHS EED) and the Health Technology Assessment Database (HTA)) were searched on 20th October 2017. The database searches were restricted to publications in English. The search in EMBASE was restricted to 1996 onwards, however MEDLINE was searched back to 1946.

Manual searches of sixteen conference proceedings were conducted for the years of 2015-2017 and publicly available information from the following HTA bodies were searched: National Institute for Health and Care Excellence (NICE), Scottish Medicines Consortium (SMC), Common Drug Review (CADTH CDR) and Pharmaceutical Benefits Advisory Committee (PBAC). In addition, the company searched the reference lists of identified systematic reviews and recent NICE Technology Appraisals for treatments of PsA.

The searches were mostly appropriate, however some weaknesses were identified by the ERG, which may have affected the comprehensiveness of the search. Appropriate electronic databases were searched to identify relevant published literature and a selection of resources were searched to find unpublished literature. However, the company did not search any trials registers to identify relevant reports of unpublished trials (ongoing and completed) of treatments for PsA. It is therefore a

possibility that any unpublished trials, particularly of comparator studies, could have been missed by the searches presented in the company submission.

The structure of the database search strategies was appropriate, consisting of terms for PsA combined with terms for the drugs used to treat PsA: tofacitinib, abatacept, adalimumab, etanercept, golimumab, infliximab, certolizumab pegol, ustekinumab, secukinumab, ixekizumab, and apremilast. The ERG notes that abatacept and ixekizumab are not relevant comparators in this appraisal. However, the ERG agrees it is appropriate to search for trials studying these treatments in PsA to be included in the systematic review and network meta-analyses. Also, the ERG noted that the biosimilar Resima (also known as CT-P13) was missing from the search strategies. The search strategy for the Cochrane Library in Table D3 was found to have missed searches for one of the comparator drugs abtacept. Therefore, any unique studies on abtacept for PsA contained in the Cochrane Library, but not present in EMBASE or MEDLINE, would not have been identified.

The search strategy for MEDLINE (Table D2) provided in the company submission was found to contain reporting errors at lines 4, 12, 20, 22, 24. These search lines were for medical subject headings that do not exist in MEDLINE. However the company provided a corrected MEDLINE strategy (in their responses to the points for clarification) to show that these search lines were searches of the “multi-purpose” (mp) field and author keywords (kw) field and not medical subject heading searches. In addition, the company clarified that the actual number of hits retrieved from the MEDLINE search was 1404 and not 1415 as originally reported at line 33 of the MEDLINE strategy (Table D2). These types of reporting errors could have been avoided by copying and pasting the search strategies from each database at the time of running the search and presenting these strategies without editing in the report. This is recommended in CRD’s guidance for undertaking reviews in health care and helps increase transparency of the searches.

The EMBASE search strategy contained a line to remove conference abstracts from the search results. Although manual searches of relevant conference proceedings were carried out by the company, these were limited to those from 2015-2017. EMBASE could have provided results of relevant conference abstracts prior to this date. It was also noted that the EMBASE strategy did not include searches of the drug trade name field (tn). Searching in this field could have improved the comprehensiveness of the EMBASE search.

4.1.2 Inclusion criteria

The inclusion criteria for the systematic review specified randomised control trials (with parallel design) of tofactinib, bDMARDs and the PDE-4 inhibitor apremilast, for the treatment of active PsA

in adults with a previous inadequate response to csDMARD therapy, which reported relevant clinical and health-related quality of life, including adverse event outcomes. The inclusion criteria were further refined to include studies of the licensed formulation of tofactinib (5mg, BD). Studies that recruited patients who suffered from other rheumatic or dermatological conditions and DMARD naïve patients were excluded. Case reports, commentaries and editorials, observational studies, and cross-sectional studies were also excluded. Only studies reported in English were eligible for inclusion. Comparators included bDMARDs, the PDE-4 inhibitor apremilast and controls including placebo, best supportive care, and any csDMARD. Studies were screened by title and abstract according to pre-defined PICOS criteria. Those that met the criteria were screened at full text. Appropriate methods were used to reduce reviewer error and bias with two blinded reviewers conducted screening of literature and any discrepancies resolved with assistance from a third reviewer.

Appropriate methods were used to extract data from the included studies. Two reviewers, blinded to each other's decisions, conducted data extraction independently, with a third reviewer involved in resolving discrepancies. Relevant data extracted from included studies are detailed in Appendix D, section D.1.6.

4.1.3 Quality assessment

Randomised control trials were assessed using the NICE Quality Appraisal checklist for quantitative interventions that assesses RCT's based on seven domains. The results of this quality assessment are presented in CS Appendix D, section D.1.7. A risk of bias assessment was also conducted assessing sequence generation, allocation concealment, baseline imbalances, blinding of participants and researchers, incomplete outcome data and selective reporting. These results are also presented in CS Appendix D; section D.1.7, along with support for judgement. The results of these assessments are given in Section 4.2.2 and Section 4.2.3 of this report.

4.1.4 Evidence synthesis

The CS focuses on two studies with distinct populations, OPAL Broaden for TNFi-naïve and OPAL Beyond for TNFi-experienced patients. The company presents the effectiveness of tofactitinib compared with the comparator treatments in forest plots in CS Appendix E. Pooled direct estimates of treatment vs placebo were presented for tofacitinib (in combination with a csDMARD) (for which results remained the same given there was only one trial per population), and the comparator treatments: adalimumab, apremilast, etanercept, infliximab, ustekinumab, golimumab, secukinumab, certolizumab pegol and ixekizumab. These analyses were conducted for the outcomes, ARC 20, 50 and 70, PASI 50, 75 and 90, PsARC and HAQ for PsARC responders and non-responders. Direct estimates pooled by drug class are also presented for the outcome PsARC.

A network meta-analysis was performed, using indirect comparisons to compare the efficacy of tofactitinib and the comparator treatments. The network meta-analysis is described in Sections 4.3 and 4.4 of this report.

4.2 Critique of trials of the technology of interest, their analysis and interpretation (and any standard meta-analyses of these)

4.2.1 Relevant trials - OPAL Broaden and OPAL Beyond

Two RCTs of tofacitinib in combination with a csDMARD were included in the CS: OPAL Broaden and OPAL Beyond. Both trials are Phase III randomised, multicentre, double-blind placebo controlled, parallel group trials, but included different populations, comparators and duration of longer term follow-up. OPAL Broaden included only TNF-inhibitor naïve patients and included a comparison with adalimumab; after the 3-month placebo-controlled phase patients were followed up on tofactinib or adalimumab to 12 months. OPAL Beyond included only TNF-experienced or intolerant patients, and did not include a comparison with adalimumab; after the 3 month placebo-controlled phase patients were followed up on tofactinib to 6 months. Details of both trials are presented in the CS – Tables 4, 5 and 6 and summarised in Table 1. After completion of these trials patients could enter a non-RCT open-label follow-up study of tofacitinib, OPAL Balance. Further details of OPAL Balance are given in Section 4.2.4.

Table 1 Summary of efficacy trials OPAL Broaden and OPAL Beyond (Adapted from CS Tables 4, 5 and 6)

Study	OPAL Broaden (2017)	OPAL Beyond (2017)
Study design	Phase 3 randomised, multicentre, 12-month, double-blind, double-dummy, active-controlled and placebo-controlled, parallel treatment group	Phase 3 randomised, multicentre, 6-month, double-blind, placebo-controlled, parallel-group
Population	Subjects with active PsA who had an IR to at least one csDMARD due to lack of efficacy or toxicity/lack of toleration and had not previously received any TNFi treatment	Subjects with active PsA who had an IR to at least one TNFi, as determined by a lack of efficacy or the occurrence of an AE that was considered by the treating physician to be related to treatment
Intervention(s)	Tofacitinib 5 mg BD (N=107) Tofacitinib 10 mg BD (N=104) Patients were required to receive a stable background dose of a single csDMARD throughout the trial	Tofacitinib 5 mg BD (N=131) Tofacitinib 10 mg BD (N=132) Patients were required to receive a stable background dose of a single csDMARD throughout the trial
Comparator(s)	Adalimumab 40 mg SC q 2 weeks (N=106) Placebo (for 3 months; N=105) At the end of the 3-month placebo-controlled period, the PBO group switched either to TOF 5 mg BD (N=52) or TOF 10 mg BD (N=53)	Placebo (for 3 months; N=131) At the end of the 3-month placebo-controlled period, the placebo group switched either to TOF 5 mg BD (N=66) or TOF 10 mg BD (N=65)
Outcomes assessed in the trials and	Primary outcomes • **ACR20 response rate at Month 3**	

relevant to the decision problem	• **ΔHAQ-DI at Month 3** Supportive analysis of primary outcomes • HAQ-DI responder analysis (≥0.35 as the cutpoint for response) at Month 3 Secondary outcomes • ACR20 response rate: Week 2, Month 6, 12 • Δ van der Heijde-mTSS, progressor rates, and non-progressor rates: Month 12 (OPAL Broaden only) • ΔACR components: Month 3 • ACR50/70 response ratea: Month 3, 6, (and 12 OPAL Broaden only) • **PASI75 response rate: Month 3,** 6, (and 12 OPAL Broaden only) • **PsARC response rate: Month 3**, 6, (and 12 OPAL Broaden only) • ΔLEI, ΔSPARCC, ΔDSS: Month 3, 6, (and 12 OPAL Broaden only) • ΔSF-36 (PF component), FACIT-F (total score): Month 3, 6, (and 12 OPAL Broaden only) (EQ-5D) Other outcomes • MDA response rate: Month 3, 6, (and 12 OPAL Broaden only) • ΔDLQI, ΔISI: Month 3, 6, (and 12 OPAL Broaden only) • ΔHAQ-DI: Month 6, (and 12 OPAL Broaden only) • ΔACR components: Month 6, (and 12 OPAL Broaden only) Post-hoc analyses used in the economic model • PASI50/90 response rate: Months 3, 6, (and 12 OPAL Broaden only) • ΔHAQ-DI conditional on PSARC response status: Month 3, 6, (and 12 OPAL Broaden only)

The inclusion criteria for both trials were: adults aged ≥18 years; diagnosis of PsA for ≥6 months; meeting the CASPAR[11] criteria at screening; active arthritis (≥3 tender/painful and ≥3 swollen joints); and active plaque psoriasis at screening and baseline. For OPAL Broaden, patients had to have demonstrated an inadequate response (lack of efficacy and/or tolerability) to ≥1 csDMARD and to have received no previous TNFi treatment; prior use of non-TNFi bDMARDs for treatment of psoriasis must have been discontinued for ≥6 months prior to the first dose of study drug. For OPAL Beyond, patients had to have demonstrated an inadequate response to ≥1 TNFi. Details of exclusion criteria, which were the same for both trials, are given in CS Table 6.

Analysis sets and statistical methods

In both trials the analysis of efficacy was based on the full analysis set (FAS) which comprised all randomized patients who received at least one dose of the randomised study drug. In OPAL Broaden this comprised all randomized patients (tofacitinib 5 mg n=107, adalimumab n=106, and placebo n=105); in OPAL Beyond it comprised all but one patient randomized to tofacitinib 5 mg (tofacitinib 5 mg n=131 and placebo n=131). It should be noted that [redacted]

[redacted]

The statistical methods used in OPAL Broaden and Beyond were similar and are summarised in Section B2.4.2 of the CS. The methods are appropriate with both trials having over 90% power to detect a 20% treatment difference and OPAL Beyond having 84% to detect a 15% treatment difference, though the treatment difference for many outcomes is much smaller than this.

Type I error was adjusted for multiple comparisons for ARC 20, change in HAQ-DI at three months and the secondary end-points PASI75, ΔLEI, ΔDSS, ΔSF-36 Physical Functioning Domain and ΔFACIT-F total score at Month 3. As requested by the ERG, the company provided additional detail of the methods used to adjust for multiple comparisons. The response stated that a gate-keeping or step-down strategy was used to protect the global type one error; specifically which step-down method was used was not clear. Three families of hierarchical testing procedure were used:

- Primary and key secondary endpoints at Month 3 (Global type I error)
- The ACR family responses (ACR20/50/70) at Month 3
- ACR20 time course (Month 3, Month 2, Month 1, Week 2)

For secondary analyses whereby steps were not taken to control for type I error, the CS states the p-values are nominal. The ERG considers the methods used to be broadly appropriate.

Missing data and withdrawals were dealt with as follows: non-responder imputation was applied to response-type/binary endpoints: ACR20, ACR50, ACR70, ΔHAQ-DI (decrease) ≥0.35, PsARC, PASI75, and MDA. No imputation was applied to missing HAQ-DI data. Missing mTSS values at Month 12 (OPAL Broaden only) were imputed via linear extrapolation.

ERG comments on design and generalisability of the trials

The ERG notes that the design of the two OPAL RCTs is appropriate to address the questions of the efficacy of tofacitinib for the treatment of active psoriatic arthritis. The study design and inclusion criteria are similar to the RCTs of already approved TNF inhibitors and other biologic DMARDs and apremilast, and the outcomes assessed are appropriate. Although the duration of the trials is 12 and 6 months respectively, unfortunately the length of the placebo-controlled period in each trial is only 3 months. However, this assessment duration, whilst limited, is in line with that used in efficacy trials of other agents in active psoriatic arthritis.

It is important to note that in all arms of the trials patients receive a csDMARD in addition to the trial therapy. Therefore the tofacitinib arm is not fully reflective of clinical practice as the licence for tofacitinib in PsA specifies concomitant therapy with MTX. This is discussed further in Section 4.2.2.2. Also of particular interest in OPAL Broaden is the comparison with adalimumab: this randomised, double-blind comparison had a 12-month follow-up, providing clear evidence for the comparison with an established TNFi. It should be noted again however, that the concomitant use of a csDMARD means the results in the adalimumab arm are not fully reflective of clinical practice, nor comparable with those from other adalimumab trials: in both contexts only a proportion of patients would take concomitant csDMARD. In addition, it should be noted that the trial was not powered to test the comparison between tofacitinib and adalimumab; this needs to be taken into consideration when interpreting any noteworthy treatment differences that do not reach statistical significance.

4.2.2 Results of OPAL Broaden

4.2.2.1 Participant flow in OPAL Broaden

Participant flow in OPAL Broaden is presented in Appendix Figure D13. In summary, 422 patients were randomised and 373 (88.4%) completed the trial (Placebo 87/105 (82.9%), tofacitinib 5 mg 96/107 (89.7); tofacitinib 10 mg 96/104 (92.3%); and adalimumab 94/106 (88.7%). Percentage discontinuations were higher in the placebo arms, though somewhat surprisingly none of the discontinuations from the 10mg placebo group were for insufficient response. Adverse events were roughly equal across all treatment arms. In their clarification response the company clarified that in the group randomised to tofacitinib 5 mg, [redacted] patients withdrew by 3 months, [redacted] month 12. [redacted].

4.2.2.2 Patient characteristics of OPAL Broaden

As the tofacitinib 10 mg dose is not licensed and is therefore not relevant to the present appraisal, results for this treatment arm were not included in the CS nor in this report. The main baseline patient characteristics are presented in CS Table 7. These were similar across the tofacitinib 5 mg, adalimumab, and placebo groups, with the exception of significant differences between groups in the mean swollen-joint count (unadjusted p=0.03 for the comparison among all four trial groups), mean Leeds Enthesitis Index (LEI) score (unadjusted p=0.02 for the comparison among all four groups), and the rate (%) of MTX use at baseline (unadjusted p=0.02 for the comparison among all four groups), which were all lower in the adalimumab group, and significant differences among trial

groups in the rate of glucocorticoid use at day 1 (unadjusted p=0.02 for the comparison of the 10 mg tofacitinib BD group with other groups), which was 27% for tofacitinib 5 mg BD, 22% for adalimumab, 17% for placebo, and 11% for tofacitinib 10 mg BD. These differences would favour adalimumab slightly.

The majority of the subjects were white (97 to 99%); the mean age ranged from 47.4 to 49.4 years and the mean duration of PsA ranged from 5.3 to 7.3 years. Out of the 318 patients, 216 (67.92%) had enthesitis and 177 (55.66%) had dactylitis. Importantly only 262 (82.39%) patients were receiving concomitant MTX. The ERG notes that almost 18% of patients were therefore not receiving tofacitinib in accordance with the product licence. An analysis of the data relating to the concomitant MTX subgroup was not presented in the CS (or the CSR).

4.2.2.3 Summary of the quality of OPAL Broaden

Table 2 Quality assessment and Risk of bias assessment (Adapted from CS Tables D16 and D17

OPAL Broaden	ERG comment	Quality Assessment (NICE checklist)	Risk of Bias
	Support		Judgement
Appropriate randomization / Sequence generation	"Randomly assigned in a 2:2:2:1:1 ratio, by means of an automated Web-based randomization system"	Yes	Low
Treatment allocation concealment	"Randomly assigned in a 2:2:2:1:1 ratio, by means of an automated Web-based randomization system"	Yes	Low
Prognostic factors balanced at study outset	"The demographic and disease characteristics of the patients at baseline were similar across groups"	No	Low
Blinded to treatment		Yes	
Blinding of participants and researchers	"Placebo was provided as oral tablets and prefilled syringes matching those of tofacitinib and adalimumab, respectively. All patients received both tablets and injections to maintain the blind."		Low
Blinding of Outcome assessment	All rheumatological and dermatological assessments were performed by qualified, trained assessors who were blinded to the patient's safety data, previous efficacy data, and treatment randomization		Low
Unexpected imbalances in dropouts		no	
Incomplete outcome data	10-30% drop-outs in all groups except one, Reasons reported. No ITT. "Efficacy analyses included all the patients who underwent randomization and received at least one dose of tofacitinib,adalimumab, or placebo"		High
Measured more outcomes than reported/selective reporting	Results reported for all key outcomes	No	Low
Appropriate analysis performed		Yes	
Overall judgement			High

The ERG agrees with the quality/risk of bias assessment results reported except for the high risk of bias assigned due to incomplete outcome data. This should not apply to those outcomes where non-response imputations were applied (response-type/binary endpoints: ACR20, ACR50, ACR70, ΔHAQ-DI (decrease) ≥0.35, PsARC, PASI75, and MDA) No imputation was applied to missing HAQ-DI data, and therefore a high risk of bias might apply, but at 3-months data were almost complete (95 to 97%) and at 6 and 12 months (tofacitinib vs adalimumab) they were 93% and 89% to 90% respectively. Modified TSS values at Month 12 were available for [redacted]; values for 6 patients were imputed via linear extrapolation, but the impact on the results was small and the risk of bias appears to be low for this outcome.

4.2.2.4 Summary of efficacy results for OPAL Broaden

The results for the key efficacy outcomes are summarised in Table 3.

Table 3 Efficacy results for OPAL Broaden (FAS) ACR 20, 50 and 90, PSARC, PASI 75 and HAQ-DI.

	Month	PBO	TOF 5 mg	ADA	TOF 5 mg vs placebo (% Difference and 95% CI) p value	ADA vs placebo (% Difference and 95% CI Nominal p value	TOF 5 mg vs ADA Nominal p value
ACR 20 Response rate, n (%)	3	35/105 (33)	54/107 (50)	55/106 (52)	17.1 (4.1, 30.2), 0.01§	18.6% (5.5, 31.7), [redacted]†	[redacted]
	6	[redacted]	[redacted]	[redacted]	_	_	[redacted]
	12		73 (68)	64 (60)	_	_	[redacted]
ACR 50 Response rate, n (%)	3	10/105 (10)	30/107 (28)	35/106 (33)	18.5% (8.3, 28.7) 0.001	23.5% (12.9, 34.1) [redacted]†	[redacted]
	6		[redacted]	[redacted]			[redacted]
	12		48/107 (45)	43/106 (41)			[redacted]
ACR 70 Response rate, n (%)	3	5/105 (5)	18/107 (17)	20/106 (19)	12.1% (3.9, 20.2) 0.004	14.1% (5.6, 22.6) [redacted]†	[redacted]
	6		[redacted]	[redacted]			[redacted]
	12		25/107 (23)	31/106 (29)			[redacted]
PSARC response rate, n (%)	3	47/105 (44.8)	55/107 (51.4)	65/106 (61.3)	6.6 -6.8, 20.1 [redacted]	16.6 3.3, 29.8 [redacted]†	[redacted]
	6		[redacted]	[redacted]			[redacted]

	12		69/107 (64.5)	69/106 (65.1)			[redacted]
PASI75 response rate, n (%)	3	12/82 (15)	35/82 (43)	30/77 (39)	28.1 14.9, 41.2 <0.001	24.3 11.0, 37.6 [redacted]†	[redacted]
	6		[redacted]	[redacted]			[redacted]
	12		46/82 (56)	43/77 (56)			[redacted]
HAQ-DI score	3						
N*		102[a]	103	101			
LS mean change from baseline		-0.18	-0.35	-0.38	-0.2 (-0.3, -0.05) 0.006§	-0.2 (-0.3, -0.1) [redacted]†	[redacted]
	6						
N*			[redacted]	[redacted]			
LS mean change from baseline			[redacted]	[redacted]			[redacted]
	12						
N*			96	94			
LS mean change from baseline (SE)			-0.54 (0.05)	-0.45 (0.05)			[redacted]

§p-value is subject to the step-down approach; †nominal p-value for comparison between adalimumab and placebo; [a]One placebo subject was excluded from the analysis (no post-baseline assessments)

PASI50 and PASI90 response at month 3 were additional outcomes examined in a post-hoc analysis conducted to inform the economic model for the UK NICE submission and are presented in CS Appendix M.

The joint primary outcomes were ACR 20 response rate and HAQ-DI score, both at 3 months. For these and all of the other outcomes in these tables, with the exception of PsARC, tofacitinib was statistically significantly more effective than placebo. It should be noted that the PSARC response in the tofacitinib 5 mg arm was similar to that for ACR 20 (51.4% and 50% respectively), but the placebo rate for PsARC was much higher than for ACR 20 (44.8% versus 33%).

Although not a primary analysis, the data and results are also presented for a comparison with adalimumab. At the 3 months for all outcomes in these tables, adalimumab was statistically significantly more effective than placebo. Comparison of tofacitinib with adalimumab at 3, 6 and 12 months shows that numerically for most outcomes adalimumab was very slightly better than tofacitinib, but for no outcome was the difference statistically significant; the trial was not powered to test such a small difference.

Results were similar for other secondary measures of disease activity at Month 3, Month 6, and Month 12 and were reported and presented in CS Appendix M.

- The MDA response rate (CS Table M6) at Month 3 in the tofacitinib 5 mg BD, adalimumab and placebo groups was 26%, 25% and 7% respectively, with [redacted] for both comparisons with placebo. For tofacitinib 5 mg vs adalimumab, [redacted] The rates were sustained up to Month 12.
- Across measures of enthesitis (LEI, SPARCC) and dactylitis (DSS) (CS Table M5) at month 3 tofacitinib 5 mg BD was numerically but not statistically superior to placebo, with responses sustained up to month 6 and month 12. The results for adalimumab were similar to those for tofacitinib except for the LEI score, for which adalimumab was statistically significantly greater than placebo and the difference for adalimumab from placebo (-0.7 (95% CI -1.2, -0.1) was numerically superior to tofacitinib from placebo (-0.4 (95% CI -0.9, 0.2).
- The results for quality of life measures were presented in CS Table M7. Although most differences were nominally statistically significant, statistical significance could not be claimed due to the hierarchical testing scheme (tofacitinib was not statistically significantly superior for LEI score). Tofacitinib 5 mg BD was numerically (SF-36 PF, FACIT-F total score) and significantly (DLQI, ISI) superior to placebo at Month 3, with responses sustained up to Month 6 and Month 12. Results were similar for adalimumab, though the difference from placebo for adalimumab was numerically lower for FACIT-F and ISI score [redacted]). It should be noted that although EQ-5D data were collected in the trial these data were not included in the CS. The ERG requested these data and they were provided in the company's clarification response. The results suggest [redacted]; no formal testing was presented.

Radiographic assessment of disease progression at 12 months is summarised in Table 4. There is no placebo comparison as the placebo controlled phase of the study stopped at 3 months. At 12 months, there was evidence of a reduction in progression in the adalimumab but not the tofactinib arm, though the treatment difference was not statistically significant; again, the trial was not powered to test such a small difference. The proportion of progressors (defined as patients with an increase in mTSS of >0.5) was low in both treatment arms.

Table 4 Radiographic progression results for OPAL Broaden (FAS)

	Month	TOF 5 mg	ADA	TOF 5 mg vs ADA Nominal p value
Change in van der Heide-mTSS (LS mean) (SE)	12	0.01 (0.07) [98]	-0.07 (0.07) [95]	[redacted]
mTSS progressor rate, n/N (%)	12	[redacted]	[redacted]	[redacted]

;[§]p-value is subject to the step-down approach; [†]nominal p-value for comparison between adalimumab and placebo; [a]One placebo subject was excluded from the analysis (no post-baseline assessments)

The ERG enquired about the data, if any collected on those patients who were randomised to placebo and then switched to active treatment at the 3-month time point. In their clarification response the company provided the results at 6 and 12 months for these patients. Overall, the results reflect those for patients randomised to tofacitinib 5 mg group and are supportive of the main analysis data, though the results for PASI75 were lower than those at 3 and 6 months in the main analysis tofacitinib group.

4.2.3 Results of OPAL Beyond

4.2.3.1 Participant flow in OPAL Beyond

Participant flow in OPAL Beyond is presented in Appendix figure D15 of the CS. In summary, 395 randomised and 345(87.3%) completed the trial (Placebo 112/131 (85.5%), Tof 5 mg 122/132 (92.4); tofacitinib 10 mg 111/132 (84.1%). Percentage discontinuations and withdrawals due to adverse events were roughly equal across all relevant treatment arms (were higher in the tofacitinib 10 mg arms). In their clarification response, the company clarified that in the group randomised to tofacitinib 5 mg, five patients withdrew by 3 months, two due to AEs, one due to inadequate response and two due to other reasons. Nine discontinued by 6 months (a further four patients (three due to AEs and one for other reasons). None of the adverse events were considered to be treatment related.

4.2.3.2 Patient characteristics of OPAL Beyond

As for OPAL Broaden, the tofacitinib 10 mg dose is not included in the CS or in this report. The main baseline patient characteristics are presented in CS Table 7 were similar across the tofacitinib 5 mg and placebo groups except that there were more female subjects in the placebo group (61%) than the tofacitinib 5 mg BD group (49%). The majority of the subjects were white (90 to 92%); the mean age ranged from 49.0 to 49.5 years; and the mean duration of PsA ranged from 9.4 to 9.6 years. Out of the 262 subjects, 176 (67.18%) had enthesitis and 129 (49.24%) had dactylitis; 199 (75.95%) of subjects were receiving concomitant MTX. This is similar to the OPAL Broaden population except that the mean duration of PsA is longer. The ERG notes that almost 24% of patients in OPAL Beyond were

not receiving tofacitinib in accordance with the product licence. An analysis of the data relating to the MTX subgroup was not presented in the CS (or the CSR).

Additional information regarding previous PsA therapies was available in the CSR and was provided in the company's clarification; these are summarized in Table 5. It should be noted that these counts of previous TNFis are irrespective of whether the patient had or had not also taken a non-TNFi biologic.

Table 5 OPAL Beyond prior drug treatments for PsA by treatment group (safety analysis set)(adapted from the Company's clarification response tables)

	[redacted]	[redacted]
[redacted]		
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]		[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]

The ERG notes that whilst all patients had been exposed to one or more TNFi, [redacted] (no patient had received just a non-TNFi b DMARD). The proportion of patients who had received just one prior TNFi was slightly lower in the tofactinib than in the placebo group [redacted] These differences would tend to favour placebo. The ERG notes that these data reveal that the majority of patients in the trial (around [redacted]) had received only one TNFi. In clinical practice, it might be expected that this figure would be lower, with tofacitinib reserved for later in the treatment pathway, raising a question over the generalisability of the results as efficacy would likely be lower in a more treatment refractory population. These data also reveal that in the trial adalimumab, etanercept and infliximab were by far the most commonly received prior bDMARDS. In clinical practice a higher proportion of ustekinumab and secukinumab might be expected given their recent approvals by NICE for PsA.

4.2.3.3 Summary of the quality of OPAL Beyond

Table 6 Quality / Risk of Bias assessment results for OPAL Beyond

OPAL Beyond	ERG comment	Quality Assessment (NICE checklist)	Risk of Bias
	Support		Judgement
Appropriate randomization / Sequence generation	"A centralized automated randomization system was used to assign patients, in a 2:2:1:1 ratio"	Yes	Low
Treatment allocation concealment	""A centralized automated randomization system was used to assign patients, in a 2:2:1:1 ratio"	Yes	Low
Prognostic factors balanced at study outset	"The demographic and disease characteristics of the patients at baseline were similar across the groups, with the exception of the mean number of tender or painful joints, for which a significant difference was seen across trial groups"	No	Unclear
Blinded to treatment		Yes	
Blinding of participants and researchers	Stated as double blinded. " The investigators, patients, and sponsor were unaware of the trial-group assignments for the duration of the trial". "Matching placebo tablets were used to maintain the blinding"		Low
Blinding of Outcome assessment	"cardiovascular events, and hepatic events were adjudicated by independent expert committees whose members were unaware of the trial-group assignments" "The investigators, patients, and sponsor were unaware of the trial-group assignments for the duration of the trial"		Low
Unexpected imbalances in dropouts		no	
Incomplete outcome data	3 groups out of 4 had 10-30% drop-outs. One group had <10% drop-outs, Reasons reported. No ITT. "Efficacy analyses included all the patients who underwent randomization and received at least one dose of tofacitinib, adalimumab, or placebo"		High
Measured more outcomes than reported/selective reporting	Results reported for all key outcomes	No	Low
Appropriate analysis performed		Yes	
Overall judgement			High

The ERG agrees with the quality / Risk of Bias assessment results reported in the CS except for the high risk of bias assigned due to incomplete outcome data. This should not apply to those outcomes where non-response imputations were applied (response-type/binary endpoints: ACR20, ACR50, ACR70, ΔHAQ-DI (decrease) ≥0.35, PsARC, PASI75, and MDA). No imputation was applied to missing HAQ-DI data, and therefore a high risk of bias might apply but at 3 months, data were available for 95% tofacitinib patients and 89% placebo.

4.2.3.4 Summary of efficacy results for OPAL Beyond

Table 7 Efficacy results for OPAL Beyond (FAS) ACR 20, 50 and 90, PSARC, PASI 75 and HAQ-DI (adapted from CS Tables 15 to19).

	Month	TOF 5 mg	PBO	TOF 5 mg vs placebo (% Difference and 95% CI) p value
ACR 20 Response rate, n (%)	3	65/131(50)	31/131 (24)	26.0 (14.7, 37.2) <0.001§
	6	78/131 (60)		
ACR 50 Response rate, n (%)	3	39/131 (30)	19/131 (15)	15.3(5.4, 25.2), 0.003
	6	50/131 (38)		
ACR 70 Response rate, n (%)	3	22/131 (17)	13/131 (10)	6.9 (-1.3, 15.1), [redacted]
	6	28/131 (21)		
PSARC response rate, n (%)	3	[redacted]	[redacted]	29.8 (18.3, 41.2), [redacted]
	6	[redacted]		
PASI75 response rate, n (%)	3	17/80 (21)	12/86 (14)	7.3 (-4.3, 18.9), [redacted]
	6	27/80 (34)		
HAQ-DI score LS mean change from baseline	3	-0.39 (N=124)	-0.14 (N=117)	-0.3 (-0.4, -0.1), <0.001§
	6	-0.44 (SE 0.05) (N=122)		

§p-value is subject to the step-down approach;

PASI50 and PASI90 response at month 3 were additional outcomes examined in a post-hoc analysis conducted to inform the economic model for the UK NICE submission and are presented in CS Appendix M.

The results in Table 7 above show that there was a statistically significant benefit of tofacitinib 5 mg over placebo for the primary outcomes (ACR 20 and HAQ-DI), and also for ACR 50 and PSARC, but not for ACR 70 or PASI 75.

Results for other secondary measures of disease activity are presented in Appendix M of the CS (Tables M14 and M16). The MDA response rate at month 3 in the tofacitinib 5 mg BD group was 23% vs 15% in the placebo group, though the difference was not statistically significant ([redacted]) . The response rate in the tofactiinib group was sustained up to Month 6. For all other of these outcomes the p values for the improvements seen with tofacitinib 5 mg BD compared with placebo were all $\leq$0.01, although for LEI score, DSS, SF-36 physical functioning score, and FACIT-F total score statistical significance could not be claimed because they were subject to a hierarchical testing scheme (because the PASI75 response rate was not significant). Responses were sustained up to

Month 6. It should be noted that, as for OPAL Broaden, although EQ-5D data were collected in the trial these data were not included in the CS but were provided in the company's clarification response. The results suggest [redacted].no formal testing presented.

The ERG also enquired about the data, if any, collected on those patients who were randomised to placebo and then switched to active treatment at the 3-month time point. In their clarification response the company provided the results at 6 months for these patients. Overall, the results reflect those of those patients randomised to tofacitinib 5 mg group and are supportive of the main analysis data.

Comparison of results from OPAL Broaden and OPAL Beyond

A comparison of the results from these two trials does not reveal a consistent pattern, i.e. there is no clear indication from the results that the Beyond population is the more refractory to treatment. Compared with OPAL Broaden the placebo response was lower in Beyond for ACR 20, but it was higher for ACR 50 and 70, and also PSARC. For PASI75 the placebo response rates in the two trials were very similar; the lack of a statistically significant effect of tofacitininb in Beyond was due to a much lower tofacitinib 5 mg arm response rate compared with that seen in Broaden 21% vs 43%). The HAQ-DI results were similar across the two trials.

Regarding withdrawals from trial therapy, the ERG requested information on the number of withdrawals and whether from OPAL Beyond or OPAL Broaden, and whether the next treatment was a csDMARD or bDMARD. This information could have indicated the position of tofacitinib in the treatment pathway. However, in their clarification response the Company confirmed that neither OPAL Beyond nor OPAL Broaden were designed to assess subsequent treatments after discontinuation of tofacitinib; the requested information was not available.

The Company stated that the drug survival rates for the relevant dose of 5 mg BD tofacitinib were very high: 90% in OPAL Broaden at 12 months, and 93% in OPAL Beyond at 6 months, and only 20 patients would have required an alternative line of treatment following tofacitinib within the study duration.

4.2.4 Relevant non-randomised evidence – OPAL Balance

One relevant non-randomised study of tofacitinib in PsA was included in the CS: OPAL Balance. OPAL Balance is an open-label extension study of the long-term safety and efficacy of patients who had previously participated in OPAL Broaden and OPAL Beyond. OPAL Balance is ongoing, with an anticipated completion date of January 2020. Details are presented in CS Appendix M 2.1. In summary, all patients in OPAL Balance received tofacitinib upon entry into the study: patients were

to receive TOF 5 mg BD for one month, after which, the dose could be increased to 10 mg BD for efficacy reasons at the investigator's discretion. Doses could be reduced back to 5 mg BD for safety reasons at the investigator's discretion. The primary outcome of OPAL Balance was incidence and severity of adverse events; and change from baseline in laboratory values. Key secondary outcomes were ACR20/50/70, HAQ-DI, PsARC, PASI75, LEI, DSS.

Clarification from the company provided indirect information on the dose of tofactinib patients entering OPAL Balance had been treated with: the trial arms are summarised in Table 8 . This information revealed that of the [redacted] patients enrolled and treated in OPAL Balance from OPAL Broaden [redacted] had been treated with TOF 5 mg, [redacted] TOF 10mg and [redacted] adalimumab. Of the [redacted] patients enrolled and treated in OPAL Balance from OPAL Beyond, [redacted] had been treated with TOF 5 mg, and [redacted] TOF 10mg.

Table 8 OPAL Balance CSR Table 14.1.1.2: Subject evaluation groups by qualifying study and overall (Subjects from OPAL Broaden)

	TOF5 BD	PBO→TOF5 BD	TOF10 BD	PBO→TOF10 BD	ADA 40mg SC Q2W	All
From OPAL Broaden						
Enrolled and treated in OPAL Balance, n (%)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
From OPAL Beyond						
Enrolled and treated in OPAL Balance, n (%)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	TOF5 BD		TOF10 BD		ADA	
	[redacted]		[redacted]		[redacted]	[redacted]

This information is not particularly useful as all patients, irrespective of the treatment in the source trial, on entering Balance initially received 5 mg dose, but increasing the dose to 10 mg was permitted. Whilst the information in Table 8 tells us that only [redacted] did not have a treatment / dose alteration at the start of this study, it does not tell us how many patients were on the 10 mg dose and therefore how representative of the licensed dose (5 mg) these data are. Further information provided in the company's clarification response [redacted]. As the 10 mg dose of tofacitnib is not licensed, there is a question over the generalisability to clinical practice of the OPAL Balance data.

Table 9 OPAL Balance Patient discontinuations by month (data from second interim analysis (25 January 2017) Information taken from Company clarification response (CCR))

OPAL Balance n=686				
	Discontinuations from CCR question A5	Table 00099.4	Table 00099.4	Table 00099.4
Assessment month	Total	Total	Due to Lack of efficacy	Due to AE
3	[redacted]	[redacted]	[redacted]	[redacted]
6	[redacted]	[redacted]	[redacted]	[redacted]
9	[redacted]	[redacted]	[redacted]	[redacted]
12	[redacted]	[redacted]	[redacted]	[redacted]
18	[redacted]	[redacted]	[redacted]	[redacted]
24	[redacted]	[redacted]	[redacted]	[redacted]
36	[redacted]	[redacted]	[redacted]	[redacted]

For the January 25, 2017 data cut, safety and efficacy data from all patients in OPAL Balance were pooled, regardless of dose, due to flexible dosing between 5 mg BD and 10 mg BD. [redacted]

Baseline values for efficacy endpoints were the same baseline values used for patients in their previous clinical trial of tofacitinib.

Results

Withdrawals from OPAL Balance are presented in **Table 9**. Withdrawals at 2 years (2.5 to 3 years since start of tofacitinib) were roughly [redacted] remained on their first TNFi. This compares with 61% remaining on first anti TNFi reported for the BSBR Register.[10]The results for the change from baseline up to Month 24 (interim data analysis up to 25 January 2017) in the pooled tofacitinib group (5 mg and 10 mg BD doses) are shown in**Table 10**. These results demonstrated that improvements in signs and symptoms of the disease and physical functioning achieved by tofacitinib treatment are generally sustained long term for those patients who remain on tofacitinib therapy. The ERG notes that the number of patients in the study reduce dramatically over the 18-month period, from 634 at month 6 to 82 at month 24, presumably due to limited follow-up in a significant number of patients. This doesn't necessarily reflect drop-outs from the study, but rather the fact that the study is ongoing. Similar improvements were demonstrated for other measures of signs and symptoms of the disease (ACR50, ACR70, and PASI75), as well as measures of enthesitis (LEI), dactylitis (DSS), and pain. The ERG noted that, even though a high proportion of patients remain on tofacitinib therapy, not all achieved an ACR 20 response. In their clarification, the company confirmed that in

OPAL Balance a lack of efficacy determined by an ACR 20 response was not a criterion for withdrawal from the study.

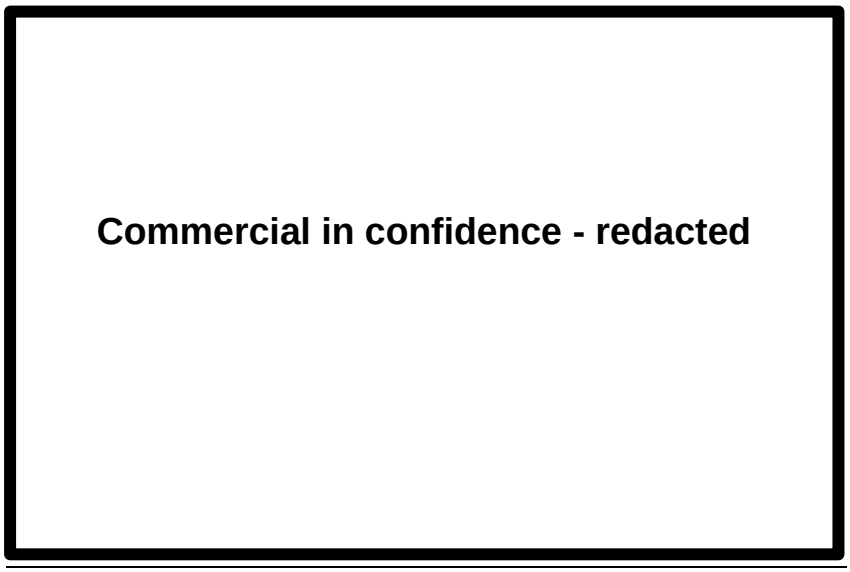
Table 10 Summary of efficacy through to Month 24 in OPAL Balance interim data analysis up to 25 January 2017 – includes TOF 5 mg and TOF 10 mg)- Includes PsARC results provided in the Company's Clarification response.

Outcome	TOF (all patients, N=686)			
Timepoint	Month 6	Month 12	Month 18	Month 24
ACR20, n/N (%)	448/634 (70.7)	422/570 (74.0)	264/341 (77.4)	55/82 (67.1)
ACR50, n/N (%)	298/633 (47.1)	284/570 (49.8)	183/342 (53.5)	41/82 (50.0)
ACR70, n/N (%)	194/636 (30.5)	183/570 (32.1)	123/341 (36.1)	22/82 (26.8)
ΔHAQ-DI, mean (SD) [N]	-0.5 (0.6) [636]	-0.5 (0.6) [571]	-0.5 (0.6) [342]	-0.6 (0.7) [81]
PSARC n/N	464/632 (73.42%)	431/566 (76.2)	271/339 (79.9)	61/82 (74.4)
PASI75 response rate, n/N1 (%)	263/433 (60.7)	250/396 (63.1)	148/242 (61.2)	40/58 (69.0)
ΔLEI, mean (SD) [N1]	-1.7 (1.8) [418]	-1.7 (1.8) [371]	-1.8 (1.8) [220]	-1.8 (1.9) [56]
ΔDSS, mean (SD) [N1]	-7.2 (7.9) [336]	-7.7 (7.8) [300]	-7.1 (7.2) [186]	-7.3 (6.6) [48]
ΔPain, mean (SD) [N1]	-26.0 (28.0) [634]	-26.8 (27.6) [570]	-29.4 (29.4) [342]	-32.6 (30.2) [81]

1= number of evaluable patients at visit. No imputation.

[redacted]

Figure 2: Change in HAQ-DI score from baseline up to Month 27 (4 April 2016 data cut) – FAS and constant tofacitinib 5 mg BD subjects only (CS Figure 7)



4.3 Evidence for impact of tofacitinib on radiographic disease progression

4.3.1 FDA Assessment – non inferiority analyses

To assess the non-inferiority (NI) of tofacitinib compared with adalimumab on radiographic outcomes the FDA developed NI margins based on two sets of data. Firstly, they conducted fixed effect (-0.63, 95% CI -0.77 to -0.48) and random-effects meta-analyses (-0.75, 95% CI -1.09 to -0.42) comparing TNFi with placebo on mean change from baseline in mTSS (modified total Sharp score) at 6 month follow up. Secondly, they used the data from the ADEPT trial on adalimumab (-1.0, 95% CI -1.60 to -0.40). Based on these data they proposed two NI margins:

- Historical data from meta-analyses of TNFi's: 0.125 to 0.375
- Historical data from adalimumab trial: 0.10 to 0.30

The upper CI for radiographic progression on tofacitinib (0.25) in OPAL Broaden is within these NI margins. However, the FDA only considered this 'borderline evidence at best' since the comparison with adalimumab was based on only one trial, and the methods used to handle missing data in the trial underestimated the standard error and therefore the width of the confidence interval (CI). In addition, there were also uncertainties regarding the constancy assumption (that the effect for the comparator observed in the OPAL Broaden reflects that of previous trials) in terms of comparability of placebo progression rates and differences in baseline characteristics.

Comparability of placebo progression rates

The FDA reviewed data on radiographic progression in psoriatic arthritis (PsA) trials. They found that placebo mean changes at 6 months ranged from 0.18 to 1.0 with mean progression greater than 0.5 in five of seven studies.

Since patients received placebo for only 3 months in OPAL Broaden, to make this comparison the FDA assumed progression at a constant rate from 3 months to 6 months. FDA concluded that the progression rates in the placebo arm of the OPAL Broaden trial were half those seen in other PsA trials historically. Similarly, mean change in erosion score for the placebo arm in OPAL Broaden was low compared with earlier studies.

Comparability of baseline characteristics in OPAL broaden with earlier trials

The FDA also compared baseline characteristics on prognostic factors such as mean baseline CRP values, baseline mTSS, erosion scores, and joint space narrowing (JSN) scores. They concluded that at baseline these values were lower in OPAL Broaden than earlier trials, potentially confounding comparisons with previous trials.

They also identified several aspects of the trial design, which further limited comparability of OPAL Broaden with previous studies. Firstly, OPAL Broaden required patients to receive a stable dose of csDMARDs. Although concomitant use of csDMARDs was not excluded in earlier trials this was not a requirement for trial inclusion (therefore some patients on placebo would have received no active treatment). Secondly, whereas in earlier trials only those who experienced an inadequate response to placebo switched to active therapy, in OPAL Broaden all patients on placebo switched to active treatment after 3 months.

FDA conclusion

The FDA concluded there is a potential effect of tofacitinib on halting radiographic progression however there is currently insufficient evidence to support this conclusion.

- Firstly, there is no evidence of difference between tofacitinib and placebo on mTSS.
- Secondly, radiographic outcomes are based on a single trial.
- Thirdly, lack of progression in the placebo arm of the OPAL Broaden is much lower than that observed in previous trials, which potentially may be explained by differences in baseline characteristics and trial design.

4.3.2 Company's analysis - Population adjusted analyses

In response to the uncertainties raised by the FDA, the company conducted population adjusted analyses based on the ADEPT trial. Differences with the FDA analyses include:

- Instead of mean difference in mTSS score for adalimumab vs placebo at 6 months, mean change from baseline mTSS score at 48 weeks was used to determine the NI margin. In addition, an NI margin was determined for rate of progression (see Table 11).

Table 11 Non-inferiority margins proposed by the FDA and the company for radiographic progression

Outcome	Source of NI margin	NI margin for upper confidence interval
mTSS	FDA: meta-analysis of TNFIs vs placebo at 6 months	0.125 to 0.375
mTSS	FDA: ADEPT trial of adalimumab vs placebo at 6 months	0.1 to 0.3
mTSS	CS: ADEPT trial of adalimumab vs placebo at 48 weeks	[redacted]
Rate of progression (change in mTSS)	CS: ADEPT trial of adalimumab vs placebo at 48 weeks	[redacted]

- The findings from OPAL Broaden were mapped to the population of the ADEPT trial adjusting for imbalances between the trial populations for potential effect modifiers and prognostic factors. Covariates assessed for inclusion in multivariable regression analyses were: baseline CRP, baseline mTSS, absence of radiographic progression at baseline, baseline erosion, baseline JSN, swollen joint count, tender joint count, use of methotrexate, RF-positive status, age in years, weight (Kg), duration of psoriatic arthritis, gender.

- In addition, covariates were centred on mean values for the ADEPT trial so that treatment differences could be interpreted within the context of the ADEPT trial population.

The population-adjusted analyses are an attempt to address the concerns raised by the FDA regarding comparability of baseline characteristics in the OPAL Broaden trial in relation to the ADEPT trial. The potential prognostic factors included in the regression model are well justified in relation to the literature.

However, there are additional potential explanations of why radiographic progression was slower in the OPAL Broaden trial other than baseline characteristics (for example, the requirement of concomitant csDMARDs for tofacitinib). Therefore, although it is possible to adjust for MTX use in

the regression models there is still potential for residual confounding due to important differences in trial design that cannot fully be adjusted for in the analyses.

Secondly, the concerns raised by the FDA regarding uncertainty associated with the non-inferiority comparisons based on a single trial remain an issue that cannot be addressed other than by further trials.

Thirdly, another source of uncertainty is length of follow up. The data for tofacitinib is based on one year follow up which is substantially shorter than data observed for TNFis. For example, the ADEPT trial provides evidence on radiographic progression up to 2.75 years and registry data provides data on radiographic progression for patients on TNFis for up to 4 years.[12]

Model selection

In univariable analyses, none of the proposed baseline covariates were associated with the treatment effect for either tofacitinib 5mg or adalimumab on mTSS at 52 weeks in the OPAL Broaden trial. Elevated CRP at baseline was associated with slightly higher odds for radiographic progression in patients receiving tofacitinib 5mg. Weight was associated with increased odds of progression in patients receiving adalimumab.

Table 12 Change in mTSS for tofacitinib 5mg vs adalimumab (adapted from table D41 in CS)

Model	Difference	p-value	95% Lower CI	95% CI Upper	AIC	Deviance
Unadjusted	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
A1:tof*(MTX)+CRP+mTSS+weight	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
A2:tof*(MTX)+CRP+mTSS+weight	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
A3: tof*(MTX+mTSS) + weight	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

tof= tofacitinib 5mg MTX=methotrexate mTSS=modified Total Sharp Score CRP=C-reactive protein

The company selected the three best fitting multivariable models for difference in mTSS for tofacitinib 5mg vs adalimumab based on the lowest AIC and deviance statistics. There were negligible differences in goodness of fit for the multivariable models compared with the unadjusted analyses (see Table 12).

Table 13 Odds of progression for tofacitinib 5mg vs adalimumab (adapted from table D42 in CS)

Model	OR	p-value	95% Lower CI	95% CI Upper	AIC	Deviance
Unadjusted	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
B1:tof*(MTX)+CRP+PSA duration+ weight+male+region	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
B2:tof*(MTX)+CRP+weight	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
B3: tof*(MTX+CRP)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

The company also selected the three best fitting multivariable models for odds of progression in tofacitinib 5mg vs adalimumab (see Table 13). As with the previous outcome, there were minor differences in goodness of fit between the multivariable and unadjusted analyses.

Comparisons with non-inferiority margins

As acknowledged in the CS the population-adjusted analyses were inconclusive as to whether reduction in radiographic progression with tofacitinib was non-inferior to adalimumab. The upper CI for the difference in mTSS and risk of progression for tofacitinib crossed both the upper and lower NI margins in unadjusted and multivariable models (see Table 14 and Table 15).

Table 14 Tofactinib 5mg on mTSS using data from ADEPT as baseline

Model	Difference	p-value	95% Lower CI	95% Upper CI
Unadjusted	[redacted]	[redacted]	[redacted]	[redacted]
Multivariable model A1:tof*(MTX)+CRP+mTSS+weight	[redacted]	[redacted]	[redacted]	[redacted]
Additional multivariable models				
A2:tof*(MTX)+mTSS+weight	[redacted]	[redacted]	[redacted]	[redacted]
A3: tof*(MTX+mTSS) + weight	[redacted]	[redacted]	[redacted]	[redacted]

tof= tofacitinib 5mg MTX=methotrexate mTSS=modified Total Sharp Score CRP=C-reactive protein

Table 15 Tofacitinib 5mg on risk of progression using date from ADEPT as baseline

Model	Risk	95% Lower CI	95% Upper CI
Unadjusted	[redacted]	[redacted]	[redacted]
Multivariable model B2	[redacted]	[redacted]	[redacted]
Multivariable model B1	[redacted]	[redacted]	[redacted]
Multivariable model B3	[redacted]	[redacted]	[redacted]

Summary

The company attempted to reduce uncertainty raised by differences in baseline characteristics between OPAL Broaden and the ADEPT trial of adalimumab in terms of population-adjusted analyses. However, there was limited evidence to show that multivariable models substantially impacted on goodness of fit. In addition, it appears there may still be differences in trial design that cannot be fully adjusted for in the analyses.

The key finding of the non-inferiority analyses is that comparisons between tofacitinib and adalimumab are currently inconclusive as upper CI's observed for tofacitinib crossed the upper and lower NI margins for both difference in mTSS and risk of progression. It cannot therefore be concluded that tofacitinib is non-inferior to adalimumab on radiographic progression outcomes.

4.4 Adverse effects of tofacitinib

Data on the adverse events associated with tofacitinib in the PsA trials (OAL Broaden, Beyond, and Balance are presented in Sections B2.11.1 to B2.11.3. The safety overview refers also to the clinical programme for tofacitinib in rheumatoid arthritis (RA).

The CS stated that over eight years of observation through the tofacitinib RA clinical programme of studies and more than 19,400 patient-years of experience have demonstrated that the rates of AEs are stable over time and are similar to bDMARDS for RA, with the exception of herpes zoster. The CS reports that in the PsA trials programme (OPAL Broaden and OPAL Beyond) tofacitinib 5 mg BD demonstrated an acceptable safety profile that is well characterised, stable, and clinically manageable. The most frequent AEs reported in the Phase III trials were nasopharyngitis, upper respiratory infection, and headache. The rate of SAEs was low across OPAL Broaden and OPAL Beyond. The types and rates of common AEs (including infections and malignancies) were generally comparable to those seen in the RA clinical programme.

[redacted]

In OPAL Broaden, where comparison with adalimumab was possible, AEs were slightly more common in the adalimumab group (see **Table 16**)

Table 16 Summary of AEs Reported up to Month 3 and Month 12 (Safety Analysis Set, All Causalities) for OPAL Broaden (adapted from CS Tables 31 and 33)

Number (%) of Subjects:	TOF 5mg, n (%)	ADA, n (%)	PBO, n (%)
To 3 months			
Subjects evaluable for AEs	107	106	105
[redacted]	[redacted]	[redacted]	[redacted]
Subjects with AEs	42 (39)	49 (46)	37 (35)
Subjects with SAEs	3 (3)	1 (1)	1 (1)
[redacted]	[redacted]	[redacted]	[redacted]
To 12 months			
Subjects evaluable for AEs	107	106	52
[redacted]	[redacted]	[redacted]	[redacted]
Subjects with AEs	71 (66)	76 (72)	36 (69)
Subjects with SAEs	8 (7)	9 (8)	3 (6)
[redacted]	[redacted]	[redacted]	[redacted]

Withdrawals due to AEs were not reported in the adverse effects section of the CS. From the trial CONSORT diagrams (CS Appendix D) and the clarification response the ERG calculated that in [redacted] and [redacted] of patients withdrew due to an adverse event in OPAL Broaden and Beyond respectively, though none of the events were considered to be treatment related. In the longer-term OPAL Balance the rate was 5.8% at 24 months.

Adverse events of special interest are summarised in the CS. These are gastrointestinal perforation and inflammatory bowel disease: tuberculosis, serious infection/herpes zoster; opportunistic infection; interstitial lung disease; cardiovascular events; and cancer. These were summarised by trial (OPAL Broaden, Beyond and Balance) but not overall; the overall totals as calculated by the ERG from the information provided are given in Table 17.

Table 17 Adverse events of special interest reported across all OPAL studies up to 36 months (ERG calculated from text in CS Appendix M)

Adverse events of special interest	N
gastrointestinal perforation and inflammatory bowel disease:	1
tuberculosis,	4 latent
serious infection	15
herpes zoster;	22
opportunistic infection;	2+ (No information from OPAL Balance)
interstitial lung disease;	0+ (No information from OPAL Balance)
cardiovascular events;	[redacted]
cancer	13

To provide long-term safety information, interim data from the long-term extension study OPAL Balance were analysed. As of January 25, 2017, no new risks or safety signals were identified in the long-term extension data from the tofacitinib PsA development programme. Types and rates of AEs (including infections and malignancies) were similar to those observed in Phase III trials and were stable over time. Recommendations on how to appropriately manage the risks associated with tofacitinib (including vaccinations and risks of serious infection) are outlined within the SmPC.

The CS also referred to a health claims database study conducted in an American cohort of PsA patients, in which the incidence of most AEs reported in tofacitinib PsA phase III studies was generally comparable with that observed in a general PsA population, with the exception of the rates of herpes zoster, which were somewhat higher in the tofacitinib cohort than in the real-world comparison cohort (Truven Marketscan Comparison Cohort).[13]

In summary, the adverse events profile of tofacitnib in PsA patients appears similar to, and no worse than that of adalimumab. The tolerability of tofacitinib is reflected in the low rate of withdrawals due to AEs. An increased risk of herpes zoster appears to be a specific AE of tofacitinib.

4.5 Critique of trials identified and included in the indirect comparison and/or multiple treatment comparison

The CS included a systematic review across the intervention of interest (tofacitinib 5 mg) and identified data on all relevant comparators (i.e. adalimumab, secukinumab, golimumab, infliximab, etanercept, apremilast, ustekinumab, certolizumab pegol). In addition, RCTs in the same populations but of interventions not included in the scope for this appraisal (abatacept and ixekizumab) were also included in the network meta-analyses this was judged to be appropriate by the ERG and discussed in

more detail below in sections 4.5.1 and 4.5.2. The methods of the review were also judged to be appropriate and are discussed in more detail above in section 4.1 and in Appendix D of the CS.

The company presented network meta-analyses (NMA) largely based on TA445, a multiple technology assessment of certolizumab pegol and secukinumab for psoriatic arthritis. The ERG compared the data included in the company analyses and the TA445 analyses and confirmed that these data overlapped in most instances.

However, as noted by the company, some inputs in the company NMA differed from TA445 where data used in TA445 were redacted in the report and unavailable in other publications. In addition, some data from other treatments were considered out of scope for TA445, but were included in the company analyses. These new inputs and their impact on findings are summarised below, for further details on studies and data included in the company NMA (see Appendix E in the CS). In addition, in response to clarifications the company provided a spreadsheet comparing findings reported in TA445 with the new data included in the company analyses.

4.5.1 bDMARD naïve population

The ixekizumab arm of the SPIRIT-P1 trial was excluded from TA445 but was included in the company's NMA (only for PASI response). Data from OPAL Broaden also contributed new data on adalimumab and tofacitinib 5mg and 10mg. While inferences obtained for the unlicenced treatments (ixekizumab and tofacitinib 10 mg) were not considered in the economic analyses, the data these studies provide may usefully inform other parameters in the NMA such as class effects, and hence the inclusions of these studies were judged to be appropriate.

Data used at 12 weeks in the FUTURE 2 (secukinumab) and RAPID-PsA (certolizumab pegol) trials were redacted in TA445. For PsARC response the company, instead used data from the mixed populations (i.e. included both bDMARD naïve and experienced patients). These new data did not make a substantial impact on the findings, although in some models this may have led to an underestimate of the effectiveness of secukinumab and certolizumab pegol (placebo adjusted and class-effect models). For PASI 75/90 16 week response data was used for secukinumab as the 12 week data was redacted in TA445. The logs odds ratios using these new data did not differ substantially from those found in TA445

4.5.2 bDMARD experienced population

New data on ixekizumab versus placebo from the SPIRIT-P2 trial and tofacitinib versus placebo from the OPAL Beyond trial were included in the company NMA analyses and these inclusions were judged to be appropriate.

Data from FUTURE 2 on secukinumab 300mg at 12 weeks were redacted in TA445. Therefore, the company's analyses included data at 16 weeks for PASI 75/90 and at 24 weeks for ACR 20/50/70. The secukinumab 300mg estimates were substantially lower for PASI 75 using the new data (TA445: 59.8% (23% to 89%); Model E1 company analyses [redacted]) but similar for PASI 90 (TA445: 36.5% (8% to 75%); Model E1 company analyses [redacted]).

4.5.3 Placebo arm of OPAL Broaden

The CS noted that the placebo response rate in the OPAL Broaden trial was the highest observed (45%) of all the included studies (Figure 3). This is consistent with TA445, which found that placebo response rates have increased over time. Therefore, the high placebo response rate is not unexpected or unique to trials of tofacitinib.

Figure 3 Placebo rates in PsA trials over time (see Figure 1 in company response to ERG question A18)

In the company response to question A18 of the ERG's points for clarification, the company suggested several potential explanations for the elevated placebo response in OPAL Broaden.

Firstly, unlike many recent trials, some of the older trials did not require patients to have failed DMARDs. This may lead to change over time in patient characteristics for those included in trials.

Secondly, it was a requirement of OPAL Broaden that all participants received a single csDMARD throughout the trial (CS, Table 5). Therefore, concomitant treatment is higher in OPAL Broaden compared with other trials (see Table 1 in response to A9 of the points for clarification letter), as although csDMARDs or methotrexate (MTX) were not excluded in previous trials, none of these previous trials required their use.

However, the ERG considered that the importance of the higher rate of concomitant csDMARDs on placebo response rates was uncertain. Most trials examined the impact of concomitant treatment but there was insufficient evidence to confirm this was an important predictor of placebo response rates. In addition, other trials with high placebo response rates such as FUTURE 2 (51% MTX), RAPID PsA (61.8%), PSUMMIT 1 (46.6% MTX), PSUMMIT2 (47.1% MTX) reported relatively low concomitant treatment in participants receiving placebo. These concomitant medication rates were similar to trials with the lowest placebo response such as GO-REVEAL (48% MTX), Mease et al 2000 (47% MTX), Genovese et al 2007 (46.9% MTX, 67.3%, other DMARDs).

Thirdly, the company observed [redacted] group as another potential explanation. As above, it is unclear to what extent this explains the higher placebo response rates in OPAL Broaden, particularly as the company did not provide estimates adjusted for geographical location.

4.5.4 Categorising placebo arms

The company's response to A18 of the ERG's points for clarification letter suggested an alternative scenario with placebo arms classified into two categories: PBO1 (older trials and apremilast trials) and PBO2 (newer trials, PSUMMIT1, RAPID-PSA, FUTURE2 and OPAL Broaden).

This categorisation partly reflects the observation in TA445 of a 'placebo creep' over time with the more recent trials reporting higher placebo response rates. However, the ERG considered there to be insufficient justification provided by the company for why apremilast trials should be categorised with the older trials, rather than those conducted from 2013 onwards.

4.6 Critique of the indirect comparison and/or multiple treatment comparison

The company submitted a number of network meta-analyses (NMA) on the effectiveness of tofacitinib (in combination with a csDMARD) compared to up to eight alternative drugs, including anti-ILs and anti-TNFs. The company presents independent analyses of 4 outcomes -- PsARC response, PASI response, HAQ conditional on PsARC response, ACR response – in 2 subpopulations -- bDMARD-naïve (of which OPAL Broaden is assumed representative), and bDMARD-experienced

(of which the OPAL Beyond trial is assumed representative). OPAL Broaden evaluates two doses of tofacitinib (10mg BD and 5mg BD) and both were included in the NMA for bDMARD-naïve. The company, however, only presented cost-effectiveness results for tofacitinib 5mg BD and hence we here omit NMA results on tofacitinib 10mg BD. This critique section will focus on the most relevant outcomes for the economic analyses (PsARC response, PASI response, HAQ conditional on PsARC response). The company presents more detailed information relevant to the NMA analyses in the appendices of the main submission:

- The results of the pivotal trials are presented in the main submission and in Appendix M.
- A description of the evidence included in the NMA, of its methods, and of the opinion of the clinical expert on the assumptions of the NMA is in Appendix D,
- The results of the NMA are presented in Appendix E.

Section 4.6 is structured as follows: We will first focus on bDMARD-naïve (section 4.6.1) and only after on bDMARD-experienced (section 4.6.2). Within each subsection, a summary of the main analyses in company's submission (including methods and results) is initially presented separately for each outcome. Note that methods of analyses differ by outcome but are similar across the two subpopulations. Hence, the general approach to modelling each outcome will be described only for the bDMARD-naïve. After summarising the company's submission, we briefly critique it. The critique will be based on comparisons with the recent TA445 that focussed on the same decision problem, and on comparison with the OPAL trials results. Finally, further detail presented in the company's submission, relevant to issues deemed important in the critique are discussed in Section 4.6.3.

4.6.1 bDMARD-naïve

This subsection summarises methods and results of the synthesis of relative treatment effects, but the company has not reported how evidence on placebo-response was considered.

4.6.1.1 Summary of main analyses in company's submission

Summary of main analyses on PsARC

The company identified 14 studies that report PsARC and organised these in a network (Figure 4, PsARC). The company only had access to the published results of Future 2 and RAPID-PsA study results (secukinumab and certolizumab pegol), which included a combination of bDMARD-naïve and bDMARD-experienced patients. In TA445 subgroup specific outcome data was used. The data on PsARC response was modelled using a standard logit model with Binomial likelihood (in line with

TA445) which expresses relative treatment effect as log odds ratios. A number of different model specifications were implemented, exploring:

- Independent treatment effects including all trial evidence (models A).
- Adjustment for differing placebo responses across trials (models B), and
- 'Class effects' alongside placebo response adjustment (models C and D). Model C considers the following classes tofacitinib 5mg BD, apremilast, TNFi, anti-IL and model D collapses TNFi and anti-IL into the same class.

All specifications were implemented using fixed and random effects across studies, respectively being identified with the numbers 1 or 2.

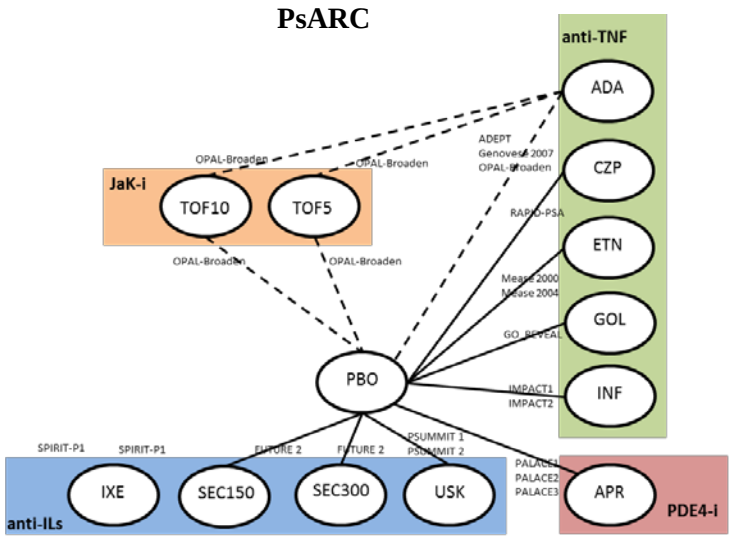
The results show that some interventions have comparable effect estimates. For ease of interpretation, in summarising results we have grouped interventions into three effectiveness levels – higher, intermediate, lower. A summary of the results across model specifications is provided below (see Table E18 in Appendix E for a detailed summary of results):

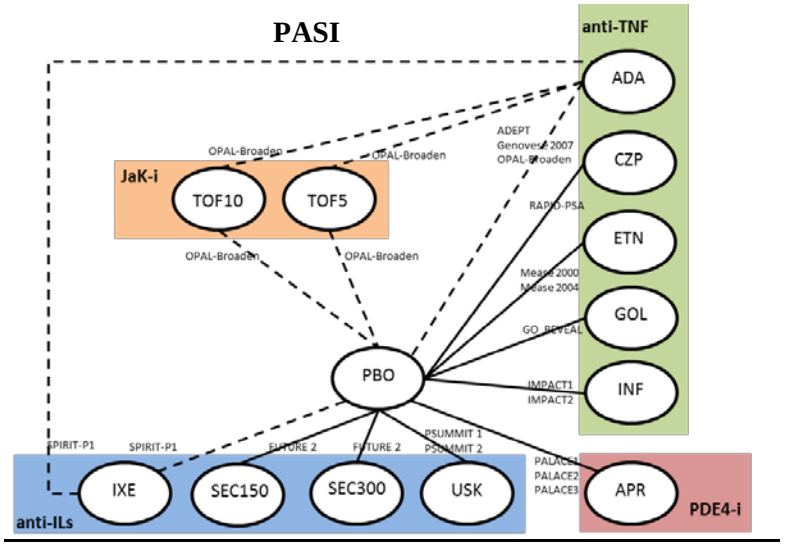
- Models A show: golimumab, infliximab and etanercept being evaluated as most effective (higher effectiveness group, here with [redacted]), followed by certolizumab, secukinumab 150, adalimumab and secukinumab 300 (intermediate effectiveness group, here with [redacted]), followed by apremilast, ustekinumab and tofacitinib 5mg BD (lower effectiveness group, here with [redacted]). The results are similar between random and fixed effects models.

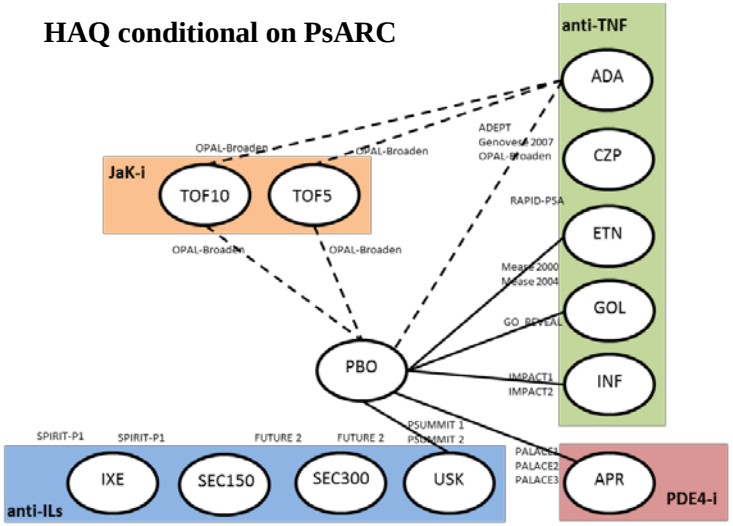
- Models B show infliximab, etanercept and golimumab ([redacted]) as most effective, followed by secukinumab 150, certolizumab, secukinumab 300 and adalimumab (intermediate effectiveness group, here with [redacted]), followed by ustekinumab, apremilast and tofacitinib 5mg BD (lower effectiveness group, here with [redacted]). Placebo-response adjustment does not significantly alter the composition of the effectiveness groups, but there are some changes in the rankings within the effectiveness groups.

- Results for models C and D do not differ significantly from models B.

Figure 4: Network diagrams for evidence on the different outcome measures







The company used goodness of fit to select from the above model specifications (see TableE18 in Appendix E). Results show that class-effects models (C and D) do not fit as well as models assuming independent effects of the different treatments, and that the placebo adjustment leads to better fitting models (B1 and B2). Within placebo-adjusted models, the random effect model, B2, has the lowest DIC. The company used model B2 as the base case for the economic model.

The results from this model are shown in Table18, where treatments are ordered according to their relative effectiveness estimates (most effective treatment is ranked 1 and the least effective is ranked 11). Effect estimates are presented using logOR against placebo (the scale in which treatment effect estimates were pooled across studies), ORs for tofacitinib 5mg BD vs other comparators, and absolute predicted PsARC response (this depends on assumptions about placebo response which were not justified in the CS).

The results highlight that all comparators were significantly better than placebo except for tofacitinib 5 mg BD (OR=[redacted]). However, when comparing across interventions, tofacitinib 5 mg BD was not significantly different to treatments in the low or intermediate effectiveness group (apremilast, ustekinumab, adalimumab, secukinumab, and certolizumab pegol), but was statistically inferior to those in the high effectiveness group (etanercept, infliximab, and golimumab). The probability of PsARC response with tofacitinib 5 mg BD was [redacted].

Table18: Main results used in the base case of company's submission (PsARC response, model B2)

	r	treat	LOR comparator vs PBO*	OR of TOF5 vs comparator	PsARC
High	1	IFX	[redacted]	[redacted]	[redacted]
	2	ETN			
	3	GOL			
Intermediate	4	SEC 150	[redacted]	[redacted]	[redacted]
	5	CZP			
	6	SEC 300			
	7	ADA			
Low	8	USK	[redacted]	[redacted]	[redacted]
	9	APR			
	10	TOF 5			
	11	PBO	[redacted]	[redacted]	[redacted]

* CI not presented in Table E18

The company also notes that OPAL Broaden is the study with the highest placebo PsARC response ([redacted]) among the included trials. For a summary of company discussion of the placebo response in OPAL Broaden and ERG critique see section 4.5.

To further explore this issue, the company submitted an additional analysis, using the specification in model A, where the placebo arm from OPAL Broaden was excluded (model A*). The manufacturer

justifies this analysis on the basis of an elevated placebo response, poor model fit in terms of residual deviance and having the support of the clinical expert that advised on the submission (see Section D.2.3 in Appendix D). This analyses returns very similar results to model A1, with the exceptions of adalimumab and tofacitinib 5mg BD, which now present better effectiveness. Specifically, in model A1 tofacitinib 5mg BD was the lowest ranking treatment (LOR of [redacted]) and in model A1* it presented better effectiveness than apremilast and ustekinumab (LOR of [redacted]).

Summary of main analyses on PASI

The evidence network used by the company is shown in Figure 4. The IMPACT trial was excluded from the NMA due to the extreme values reported in the trial (PASI 50 response was 0% for placebo and 100% for IFX). Ixekizumab was not NICE approved in the UK for PsA at the time of company's review; however the phase III study SPIRIT P1 had been published and was included in the network.

The NMA estimated the probability of PASI response at different thresholds (50/75/90) within a multinomial probit model. The single model included all categories of PASI and evaluated a single effect estimate for each treatment (expressed as a probit) that is then used to obtain probabilities of achieving PASI 50, PASI 75 and PASI 90. The company considered two alternative model specifications:

- Model E: Independent treatment effect and no placebo-response adjustment, and
- Model F: Independent treatment effect and placebo-response adjustment.

The results show that:

- Model E2 (Table E31, Appendix E and Table 2 below) identifies infliximab and ixekizumab as most effective (highest effectiveness group), followed by secukinumab and golimumab, (intermediate/high effectiveness group), followed by adalimumab and ustekinumab (intermediate/low effectiveness group), and lastly tofacitinib 5mg BD, certolizumab pegol, etanercept, apremilast (lowest effectiveness group). Results for model E1 (Table E29, Appendix E) only differ for ustekinumab, which had an effect estimate closer to adalimumab. Note that in Table 2 we omit results on ixekizumab as this is not a comparator in the submission.
- Model F does not differ from E1 indicating no effect of placebo-response adjustment.

Model selection used DIC as a goodness of fit criterion. The company found that the placebo-response adjusted FE model fitted the data as well as unadjusted FE models. The random effect model (E2), implemented only without placebo-adjustment, was used as the base case as DIC was significantly lower for this model.

The results of the base case model (E2) showed that tofacitinib 5 mg BD was not significantly different from placebo, nor from its comparators (see Table 19 and Table E31 in Appendix E). Tofacitinib 5 mg BD was estimated to have a [redacted] probability for a PASI 50 response , [redacted] for a PASI 75 response, and [redacted] for PASI 90 response

Table 19: Main results used in the base case of company's submission (PASI) (adapted from Table E31 in CS)

PASI Base case model (E2)						
			probit	PASI50	PASI75	PASI90
H – i – d	1	IFX	[redacted]	[redacted]	[redacted]	[redacted]
Inter medi ate to	2	SEC 300	[redacted]	[redacted]	[redacted]	[redacted]
	3	GOL	[redacted]	[redacted]	[redacted]	[redacted]
	4	SEC 150	[redacted]	[redacted]	[redacted]	[redacted]
Inter medi ate to	5	USK	[redacted]	[redacted]	[redacted]	[redacted]
	6	ADA	[redacted]	[redacted]	[redacted]	[redacted]
	7	TOF 5	[redacted]	[redacted]	[redacted]	[redacted]
Low	8	CZP	[redacted]	[redacted]	[redacted]	[redacted]
	9	ETN	[redacted]	[redacted]	[redacted]	[redacted]
	10	APR	[redacted]	[redacted]	[redacted]	[redacted]
PBO	11	PBO	[redacted]	[redacted]	[redacted]	[redacted]

Summary of main analyses on HAQ change conditional on PsARC response

The network of evidence for HAQ change conditional on PsARC response used by the company is shown in Figure 4. The analyses did not include Future 2 and RAPID-PsA, as the bDMARD-naïve data were redacted in TA445 and were not available in the primary publications. Hence, no results for certolizumab pegol and secukinumab for HAQ-DI could be presented in the submission.

Two alternative Normal models were used for HAQ-DI conditional on PsARC response status.

- Both G and H model the difference between placebo responders, treated responders and treated non-responders all in relation to placebo non-responders (approach used in TA445). Model G considers independent treatment effects while H evaluates class effects (classes: tofacitinib 5mg BD, apremilast, TNFi, anti-IL)
- Model K is an alternative model to the above, where data from the PsARC responder subgroup are analysed separately from the data for the PsARC non-responders. The common baseline is change in HAQ-DI for placebo responders in the PsARC responder analyses, and change in HAQ-DI for placebo non-responders in the PsARC non-responders analyses. The model adjusts the trial variance to account for multi-arm studies and the manufacturer hypothesises that a RE model would take a better account of heterogeneity.
- Placebo-adjusted models were not undertaken (in line with TA445).

In response to ERG request for clarifications, the company submitted more detailed results, corrected NMA estimates for HAQ-DI change for responders in the bDMARD-naïve population. The updated values are shown below in Table 20.

Infliximab and ETA are associated with the highest HAQ reductions in PsARC responders across all models. Of the remainder, ustekinumab, adalimumab, tofacitinib 5mg BD, and golimumab show similar results for PsARC responders, but tofacitinib 5mg BD shows much higher effects than others on HAQ for non-responders (comparable to infliximab and ETN)

Table 20: Main results used in the base case of company's submission (HAQ conditional on PsARC response, model K2) -- Corrected

	r		Predicted HAQ change	
			*Responders **	*Non-responders*
High	**1**	IFX	[redacted]	[redacted]
	2	ETN		
Intermediate	**3**	USK		
	4	ADA		
	5	TOF 5		
Low	**6**	GOL		
	7	APR		
PBO	**8**	PBO		

*results corrected in clarification

4.6.2 BDMARD-experienced

4.6.2.1 Summary of main analyses in company's submission

Summary of main analyses on PsARC

Data from 2 studies were included in the network (see Figure 5). Only model A1 was implemented (independent treatment effects, no placebo-response adjustment, see section 4.6.1). The results from the model are shown in Table 21, where tofacitinib 5mg BD is estimated to have a PsARC response very similar to ustekinumab.

Summary of main analyses on PASI

The company analyses on PASI included new evidence from the TOF comparison from OPAL Beyond, IXE from SPIRIT-P2, and ABA from ASTRAEA in addition to that used in TA445. Data from FUTURE 2 on secukinumab 300mg at 12 weeks were redacted in TA445. Therefore, the company's analyses included data at 16 weeks for PASI 75/90 and at 24 weeks for ACR 20/50/70.

Only model E1 was implemented, with and without 24-week data (the latter excludes the comparison with IXE). The results of the NMA model (with 24-week data) is shown in Table 21. Ustekinumab and secukinumab show best PASI responses, followed by IXE. Tofacitinib 5mg BD had a PASI response slightly higher but not significantly different from placebo. ABA shows response levels similar to placebo. The exclusion of 24-week data does not alter results significantly. The secukinumab 300mg estimates were substantially lower for PASI 75 using the new data (TA445: 59.8% (23% to 89%); Model E1 company analyses [redacted]) but similar for PASI 90 (TA445: 36.5% (8% to 75%); Model E1 company analyses [redacted]).

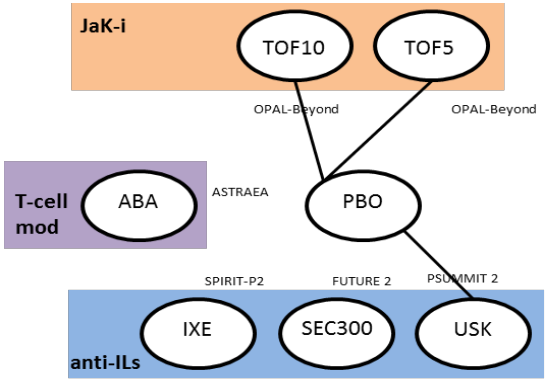
Summary of main analyses on HAQ change conditional on PsARC response

The manufacturer implemented models G and K, which were both fixed effects. Contrary to the bDMARD naïve population, in experienced patients the manufacturer chose model G for the base case and K for sensitivity analyses.

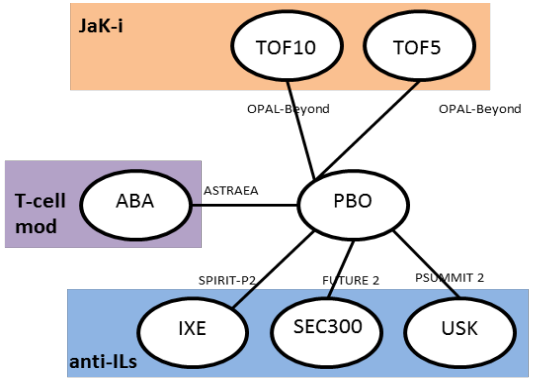
The results from model G are shown in Table 21. Results show that model G (chosen for the base case) evaluates tofacitinib 5mg BD to have higher HAQ changes than ustekinumab in both responders and non-responders, while model K presents ustekinumab as having the highest HAQ improvement in responders.

Figure 5: Network diagrams for DMARD experienced population

PsARC



PASI



HAQ conditional on PsARC

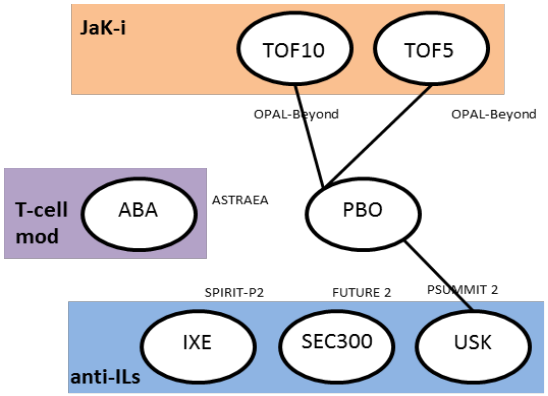


Table 21 Results from NMA in DMARD experienced population

PsARC Base case model (A1)

Rank	treat	PsARC
1	USK	[redacted]
2	TOF5	[redacted]
3	TOF10	[redacted]
4	PBO	[redacted]

PASI Base case model (E1)

		PASI50	PASI75	PASI90
1	USK	[redacted]	[redacted]	[redacted]
2	SEC 300	[redacted]	[redacted]	[redacted]
3	IXE 80 Q2W	[redacted]	[redacted]	[redacted]
4	IXE 80 Q4W	[redacted]	[redacted]	[redacted]
5	TOF 10	[redacted]	[redacted]	[redacted]
6	TOF 5	[redacted]	[redacted]	[redacted]
7	PBO	[redacted]	[redacted]	[redacted]
8	ABA	[redacted]	[redacted]	[redacted]

HAQ conditional on PsARC (Model G1)

		Predicted HAQ change	
		Responders	*Non-responders*
1	TOF10	[redacted]	[redacted]
2	TOF5	[redacted]	[redacted]
3	USK	[redacted]	[redacted]
4	PBO	[redacted]	[redacted]

4.6.3 Critique of NMA and outstanding issues

4.6.3.1 Critique of analyses implemented for bDMARD naïve population

Across all analyses of PsARC response, presented by the company, tofacitinib 5mg BD is consistently in the lower effectiveness group, which also includes apremilast. Results vary slightly across specifications in how similar its effectiveness is in relation to apremilast: e.g. in model A the LOR for apremilast is [redacted] and for tofacitinib 5mg BD is [redacted] and in model B2 (base case) apremilast's LOR is [redacted] and tofacitinib 5mg BD is [redacted]

The evidence network and data included in the company NMAs substantially overlap with TA445 (see Figure 4 for new evidence since TA445 illustrated with dashed lines).

PsARC response

The range of model specifications tested in the company analyses of PsARC outcomes was similar to TA445. However, the company's NMA results differed from those obtained in TA445:

- Results of the independent models (not adjusted for placebo response) are very similar except for adalimumab which was found to be more effective in TA445 (LOR= 1.352) than in the company submission (LOR= [redacted]).
- The results of the placebo-adjusted models (B1 and B2) differ substantially. In TA445, placebo-response adjustment had a pronounced impact on the rankings: secukinumab became most effective with a LOR of 2.1. Etanercept, infliximab and certolizumab pegol were of similar effectiveness (but LOR values reduced to below 2). Golimumab moved down in the ranking to LOR values around 1.6. LORs for ustekinumab and adalimumab were close to, but above, 1. Apremilast was still the least effective (LOR of 0.765).
- The AG in TA445 also explored placebo-response adjusted models with class effects. However, although the company include similar models data were not used to inform the cost-effectiveness analyses
- The AG in TA445 concluded that without any clear rationale for the placebo effect, the results of the placebo-response adjusted model should be interpreted with caution. The model with independent treatment effects was hence used in the base case in TA445, and the best fitting model including placebo-response adjustment and class effect was used in sensitivity analysis.
- OPAL Broaden showed a much higher PsARC placebo response (of 44.8%) than that modelled.

PASI response

TA445 applied models equivalent to E1 and F1, but random effects models were not evaluated. The results were relatively similar to the company's except for secukinumab and adalimumab. In TA445 secukinumab and adalimumab were estimated to have higher PASI responses.

As with PsARC, OPAL Broaden showed a higher placebo response on PASI (of [redacted] respectively for PASI50, PASI75 and PASI90) than that modelled. The model found adalimumab response was similar to tofacitinib 5mg BD; the trial shows, however, that while this holds for PASI 50 ([redacted] for tofacitinib 5mg BD and [redacted] for adalimumab), PASI75 and PASI90 show better results for tofacitinib 5mg BD ([redacted] for PASI75 and [redacted] for PASI90).

HAQ conditional on PsARC response

Model specifications and findings of the company analyses (model G) were similar to TA445 for HAQ changes conditional on PsARC. Predictions from model G were also consistent with the results from OPAL Broaden, including for placebo. However, there are significant differences in predictions from model K particularly in what concerns responders to PsARC.

4.6.3.2 Critique of analyses implemented for bDMARD experienced population

The PsARC response rates from the company analyses for ustekinumab were similar to those in TA445, but TA445 was able to include data for secukinumab, which showed higher effectiveness than ustekinumab. OPAL Beyond showed a similar placebo response (of [redacted]) and tofacitinib 5mg BD response (of [redacted]) to that modelled.

TA445 found lower placebo response rates for PASI (8.8% to PASI 50), and higher responses to secukinumab 300 than ustekinumab (PASI 50 of, respectively, 87.5% and 62.8%).

OPAL Beyond had a higher placebo PASI responses rate (of 26.7%, 15% and 10%, respectively for PASI50, PASI75 and PASI90) than those modelled. Responses observed in the trial for TOF are 45%, 21% and 13.75% respectively for PASI50, PASI75 and PASI90.

[redacted]. The predictions for model G are slightly closer to trial results. HAQ changes in non-responders were low and very similar in the trial.

4.6.3.3 Outstanding issues

The ERG identified no significant issues with analyses relating to the bDMARD-experienced population. There are two outstanding issues on the evidence synthesis for the bDMARD-naïve population. The first issue is of key importance, concerning the validity of the placebo-response adjusted models for the estimation of treatment effects over PsARC response on the bDMARD-naïve population. This is be explored in the next section. The second outstanding issue is the level of placebo-response for PsARC and PASI response outcomes. The manufacturer has not identified the assumptions underlying the placebo-response assumed in the models. Typically, placebo response rates are informed by synthesising data from the literature but it is not clear whether this is the case in the company analyses. However, given the values used are similar to those in TA445, this issue will not be explored further.

4.7 Additional work on clinical effectiveness undertaken by the ERG

This section will focus on two aspects of the submission on the bDMARD-naïve population:

- A correction on the PsARC models with adjustment for placebo-response (models B)
- Revisiting model selection following the model correction (models C and D).

4.7.1 Correction of placebo-response adjusted models for PsARC

Given the disparities found in the placebo-adjusted models between the company's submission and TA445, the company was asked, in response to ERG requests for clarification, to justify the differences and explore why the placebo arm of the OPAL Broaden trial did not fit well in the NMA model with placebo adjustment for PsARC. In response to clarifications two additional analyses were submitted by the company:

1. Placebo comparator arms were split into two separate comparators:
   This new analysis splits the placebo arms into two: PBO 1 (older trials and apremilast trials) and PBO 2 (newer trials, PSUMMIT, RAPID-PsA, FUTURE 2 and OPAL Broaden). The company argued the higher placebo response in newer trials might reflect a difference in previous and/or concomitant treatments between newer and older trials (except for apremilast, which has a similar placebo response to the older trials). For further discussion of differences between newer and older trials, please see section 4.5 above. The results of this new analysis (detailed in Table 2 in response to clarification document) indicate that PBO 1 had lower odds of PsARC response compared to PBO 2. All treatments hence had lower OR vs. PBO 2 than with PBO1. The model specification means that the rankings are retained between comparisons to PBO 1 and PBO 2 (no placebo adjustment). The ORs for tofacitinib, certolizumab, secukinumab, and ustekinumab when compared to PBO 2 are a better match to the trial data placebo comparisons (OPAL Broaden, RAPID-PSA, FUTURE 2, PSUMMIT1 and 2). The ORs for the TNFis etanercept, infliximab, and golimumab when compared to PBO 1 were a better match to the placebo comparisons reported in the trials (Mease 2000, 2004, IMPACT1 and 2, GO-REVEAL). In this analysis tofacitinib 5mg BD was more effective than apremilast.
2. Placebo adjustment was allowed to differ by treatment: a placebo-adjusted model specification was used, but instead of assuming a common placebo effect across treatments, the coefficient beta was allowed to vary by treatment, with all betas drawn from a common random-effects distribution. (Results in Sheet A18 in the Excel workbook that accompanies the response to clarification). This model returned different rankings to all previous models, and some nonsensical results, with apremilast evaluated as second most effective treatment.

In the clarification questions, the company was also asked to provide all files required to run the NMA models in WinBUGS (including data, model, and initial values for every chain). The ERG checked the models and found that placebo-response adjusted models were incorrectly implemented (see appendix A). This means that results presented in the main submission for models B, C and D, and for the two analyses described above, are thus incorrect.

The company's base case, model B2, was corrected by the ERG (based on 100,000 iterations with a thin of 15 from 3 independent chains after a burn-in of 50,000), and results are shown in Table **22**. The treatment effects are interpreted as the effects for patients with a baseline probability of PsARC of [redacted] (logit probability of [redacted]). The model estimates a credible region for the interaction term B far from zero, suggesting a strong interaction effect between the baseline risk and the treatment effects.

Table 22: Main results used in the base case of company's submission (PsARC response, model B2) -- Corrected

	r	treat	LOR, comparator vs PBO*	OR of TOF5 vs comparator	PsARC response
High	1	ETN	[redacted]	[redacted]	[redacted]
	2	IFX	[redacted]	[redacted]	[redacted]
Intermediate high	3	SEC 150	[redacted]	[redacted]	[redacted]
	4	GOL	[redacted]	[redacted]	[redacted]
	5	CZP	[redacted]	[redacted]	[redacted]
	6	SEC 300	[redacted]	[redacted]	[redacted]
Intermediate low	7	ADA	[redacted]	[redacted]	[redacted]
	8	USK	[redacted]	[redacted]	[redacted]
low	9	TOF 5	[redacted]	[redacted]	[redacted]
	10	APR	[redacted]	[redacted]	[redacted]
	11	PBO	[redacted]	[redacted]	[redacted]
		B	[redacted]		
		sd	[redacted]		
		sumdev	[redacted]		
		DIC	[redacted]		
		dev[13,1]	[redacted]		

* CI not presented in Table E18

Whilst only multiple studies on the same treatment and with placebo comparison contribute to estimating the placebo-response adjustment coefficient, B, the assumption of a common regression term allows this to be assumed valid in comparisons which only have one trial. This means the change in the rankings is expected (in relation to a model without placebo-response adjustment) and this also affects treatments that have only been trialled once.

Also note that the Table reports the model fit to the OPAL Broaden placebo arm -- dev[13,1] which shows residual deviance for this data point is substantially lower ([redacted] compared with [redacted] in company analyses) and implies a good fit between the data and the model.

The corrected base case model shows:

- etanercept and infliximab are the most effective drugs (higher effectiveness group, here with [redacted]), followed by secukinumab, golimumab and certolizumab (intermediate/high

effectiveness group, here with [redacted]), followed by adalimumab and ustekinumab (intermediate/low effectiveness group, here with [redacted]), and lastly tofacitinib 5mg BD and apremilast (lower effectiveness group, here with [redacted]).

The results highlight that all comparators were significantly better than placebo including tofacitinib 5 mg BD. When comparing across interventions tofacitinib 5 mg BD was not significantly different to any other treatment. The probability of PsARC response with tofacitinib 5 mg BD was [redacted].

Comparison with TA445

There is a noticeable difference in the magnitude of the coefficient on placebo-response when compared to TA445 (-1.4 in TA445 vs. -[redacted] in the CS), which explains the less pronounced effect of placebo-response adjustment on treatment rankings. This is due to the inclusion of OPAL Broaden, a study that includes a pairwise comparison between adalimumab and placebo and therefore informs the PBO effect (together with Genovese and ADEPT). [redacted] plots the crude data from the trials (log odds of placebo response on the x-axis and the log odds ratio for the intervention arm on the y-axis). Each dot in the plot represents pairwise comparisons from each study. The red dots show evidence on adalimumab vs placebo, with the far right dot representing the data for OPAL Broaden.

Commercial in confidence - redacted

The trend lines in the figure show the information that contributes to the placebo effect and in red the subset of adalimumab trials. The slope of the red trend line hence represents the information conveyed in the ADA studies on the coefficient for the meta-regression. OPAL Broaden conveys information that complements, and does not contradict, the remaining adalimumab trials (Genovese and ADEPT) regarding the placebo effect coefficient. This information should therefore not be dismissed.

4.7.2 Revisiting model selection for placebo-response adjusted models for PsARC

In this subsection, we implement all model specifications submitted by the manufacturer in order to revisit model selection after the correction to placebo-response adjusted models. The corrected inferences are presented below (Table 23), alongside goodness of fit statistics.

Table 23 : Results of a range of NMA models (PsARC response) – Corrected

model	A1	A2	B1c	B2c	C1c	D1c	ERG
ADA							
APR							
ETN							
IFX							
USK							
GOL							
TOF 5							
TOF 10							
SEC 150							
SEC 300							
CZP							
B							
SD							
Class: APR							
Class: TOF5, TOF10							
Class: TNFi							
Class: TNFi and anti-IL							
Class: Anti-IL							
Class: TOF5							
Class: TOF10							
precclass							
sumdev							
DIC							

*TNFi: ADA, ETN, IFX, GOL, CZP; Anti-IL: SEC, USK

Results show that:

- placebo-adjustment improves model fit. There is also strong evidence for the impact of placebo-response on effectiveness as its coefficient is statistically significant.
- placebo-adjustment may account for some of the heterogeneity across trials, and hence the fixed effect model (B1) now presents a marginally lower DIC than the random effects model (B2).
- Both class effect models proposed by the company (C and D) fit the data well, and provide better fit to the data than the independent treatment effect models. Model D fits the data as well as C but is most parsimonious. Note, however, that both C and D include TOF5 and TOF10 in the same class. Therefore, the effectiveness of TOF 5 is increased as information is shared across the two doses.
- The ERG extended model D to separate TOF5 and TOF10, whilst keeping all other aspects of the model the same as the company analyses. This model fitted the data as well as the other class effect models tested, but results in the lowest residual deviance and the precision for the class effect is increased.

Whilst the placebo-response adjusted models fit best to the data, the rationale for the differences in placebo-response across trials is not clear and therefore, as highlighted in TA445, the results of the placebo-response adjusted model should be interpreted with caution. We will therefore explore the use of both the independent treatment effects (A2), and of the class effect model proposed by the ERG (placebo-response adjusted class effect model) in Section 6. More detailed summaries of these two models are presented in Table 24.

Table 24: Additional summaries on preferred models for analyses (models A2 and ERG model)

		A2		ERG model	
r	treat	OR of TOF5 vs treat	PsARC treat	OR of TOF5 vs treat	PsARC treat
1	GOL	[redacted]	[redacted]	[redacted]	[redacted]
2	IFX				
3	ETN				
4	CZP				
5	SEC 150				
6	SEC 300				
7	ADA				
8	APR				
9	USK				
10	TOF 5				
11	PBO				

4.8 Conclusions of the clinical effectiveness section

Clinical effectiveness of tofacitinib

The clinical effectiveness of tofacitinib was informed by two good quality RCTs; one for TNFi naïve (OPAL Broaden) and one for TNFi experienced patients (OPAL Beyond). There was also long-term open label follow-up (OPAL Balance).

The trials demonstrated that compared with placebo tofacitib has some degree of efficacy across a range of outcomes in both TNFi naïve and TNFi experienced patients. There were no statistically significant differences between tofacitinib 5mg and adalimumab on radiographic outcomes but OPAL Broaden was not powered to test for non-inferiority.

Non-inferiority of tofacitinib compared with adalimumab on radiographic outcomes

Population adjusted analyses were also conducted to compare tofacitinib and adalimumab, using data from the ADEPT trial as baseline. Findings were inconclusive as the upper confidence interval crossed both the upper and lower NI margins in unadjusted and multivariable models. In addition, there is only data comparing tofacitinib and adalimumab up to 52 weeks and therefore longer term data on the effectiveness of tofacitinib is lacking. Therefore, concurring with the FDA conclusions, there is currently insufficient evidence to support the assumption that tofacitinib halts radiographic progression.

Generalisability

The ERG identified some issues regarding the generalisability of the trials to clinical practice:

- A significant proportion of patients in each RCT (18% and 24%) were treated in combination with sulfasalazine and leflunomide, whereas the marketing authorisation is for tofacitinib in combination with methotrexate (MXT) only.
- The adalimumab comparator in OPAL Broaden was in combination with a csDMARD. This is not reflective of adalimumab in clinical practice or in other trials: usually only a proportion of patients would use adalimumab concomitantly with a csdMARD.
- [redacted]In OPAL Beyond the number and nature of previous TNFis might not reflect how tofacitnib will be used in current practice.
- Treatment with tofacitinib is long-term but the placebo controlled phase was limited to only 3 months.

Network meta-analyses

The data and network meta-analyses (NMA) models used in the company analyses were similar to TA445, a recent multiple technology appraisal. There were two corrections made to the CS: one on HAQ changes conditional on PsARC response (detected by the manufacturer at clarification stage) and another on the placebo adjusted NMAs for PsARC (detected by the ERG).

The final NMA analyses showed that tofacitinib 5mg was consistently ranked among the least effective of the treatments for PsARC, at a similar level of effectiveness to apremilast. Whereas for PASI response and HAQ-DI conditional on PsARC response, tofacitinib 5 mg was associated with level of effectiveness more similar to adalimumab (although uncertainty over the magnitude of effect for tofacitinib is higher than for adalimumab).

The NMA on PsARC response explored an adjustment for the differing placebo response rates seen across trials (as in TA445). The best fitting model used such an adjustment, together with class effects (ERG model). However, the rationale for the differences in placebo response observed across trials is not clear, and hence the independent treatment effects (A2) was also used in the economic model.

Adverse events

The adverse events profile of tofacitnib in PsA patients appears similar to, and no worse than that of adalimumab. The tolerability of tofacitinib is reflected in the low rate of withdrawals due to AEs. An increased risk of herpes zoster appears to be a specific AE of tofacitinib.

5 Cost Effectiveness

This section focuses on the economic evidence submitted by the company and the additional information provided in response to the ERG points for clarification. The submission was subject to a critical review on the basis of the company's report and by direct examination of the electronic version of the model. Section 6 presents additional work undertaken by the ERG to address any errors, further explore key assumptions and possible limitations.

The company's economic submission included:

• A description of the systematic literature review conducted to identify published evidence on the cost-effectiveness of tofacitinib for PsA (CS, Section B.3.1.1) with a complete description of the search strategy in a separate appendix (CS, Appendix G).

• A report on the de novo economic evaluation by the company. The report described the patient population, model structure and treatment pathway (CS, Section B.3.2), the clinical parameters and variables (CS, Section B.3.3), measurement and valuation of health effects (CS, Section B.3.4), cost and healthcare resource use identification, measurement and valuation (CS, Section B.3.5), a summary of the base-case analysis inputs and assumptions (CS, Section B.3.6), the cost-effectiveness results for the base-case and sensitivity analyses (CS, Section B.3.7 and B.3.8) .

• An electronic copy of the company's economic model developed in Microsoft Excel.

In response to a number of points for clarification raised by the ERG, the company submitted:

- A descriptive reply to the ERG's points for clarification, as well as appendices with additional data requested by the ERG.
- An updated version of the company's electronic model incorporating;
    - Corrections to a data entry error in the HAQ-DI NMA results
    - Modifications to the response rate reported in the NMA that were inconsistent between the CS and the economic model
    - Flexibility to specify a separate withdrawal rate for tofacitinib

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

5.1.1 Searches

Cost-effectiveness searches

The search strategies used by the company to identify 1) relevant economic evaluations of tofacitinib and other treatments for PsA and 2) relevant studies of resource use and costs associated with the management of PsA in the UK were presented in full detail in Appendix G.

The following electronic databases were searched on 20th October 2017: Cochrane Library (including the Cochrane Database of Systematic Reviews (CDSR), the Database of Abstracts of Reviews of Effects (DARE), the Cochrane Central Register of Controlled Trials (CENTRAL), NHS Economic Evaluation Database (NHS EED) and the Health Technology Assessment Database (HTA)) and EconLit. In addition, MEDLINE, MEDLINE In-Process and EMBASE were searched on 13th November 2017 with a limit applied to restrict retrieval to English language studies. EMBASE was searched from 1996 onwards and MEDLINE was searched back to 1946.

Manual searches of the abstracts of sixteen conference proceedings were conducted for the years of 2015-2017 and publicly available information from the following HTA bodies were searched for any previous, relevant HTA submissions: National Institute for Health and Care Excellence (NICE), Scottish Medicines Consortium (SMC), Common Drug Review (CADTH CDR) and Pharmaceutical Benefits Advisory Committee (PBAC).

In addition, the company searched the reference lists of any identified systematic reviews and included references identified from the clinical effectiveness searches which met the economic inclusion criteria.

Appropriate sources of literature were searched to identify both published and unpublished studies for the cost-effectiveness systematic review and to identify studies of cost and resource use in the management of PsA. The search strategy for EconLit was missing from the company submission, however was provided by the company in their responses to the questions for clarification.

The reporting of the number of hits in the economic PRISMA flow diagram (page 23, Appendix G) was unclear in the company submission. The number of hits from MEDLINE and EMBASE was queried by the ERG, as the numbers did not match those presented in the final results of the search tables (Table G1 EMBASE and Table G2 MEDLINE, pages 5-14, Appendix G). The company replied in their responses to the points for clarification that this was due to additional economic studies found for the review from the clinical effectiveness searches. This seems reasonable but could have been presented more clearly in the PRISMA flow diagram.

The structure of the database search strategies was appropriate, however, the ERG noted that the biosimilar Resima (also known as CT-P13) was missing from the search strategies. The search strategy for the Cochrane Library in Table G3 was found to have missed searches for one of the comparator drugs abatacept. In addition, searches for abatacept and adalimumab were missing from the EconLit strategy. Therefore it is a possibility that relevant economic studies of abatacept or adalimumab for the treatment of PsA would not have been identified by the search strategies presented in the submission.

As with the clinical effectiveness searches, the EMBASE search strategy contained a line to remove conference abstracts from the search results. Although manual searches of relevant conference proceedings were carried out by the company, these were limited to those from 2015-2017. EMBASE could have provided results of relevant conference abstracts prior to this date. It was also noted that the EMBASE strategy did not include searches of the drug trade name field (tn). Searching in this additional field could have improved the comprehensiveness of the EMBASE search.

Health-related quality-of-life searches

The search strategies used by the company to identify health-related quality of life studies were described in full detail in Apppendix H.

The electronic databases MEDLINE (including MEDLINE, Epub Ahead of Print, In-Process & Other Non-Indexed Citations and MEDLINE Daily), EMBASE and the Cochrane Library (including the Cochrane Database of Systematic Reviews (CDSR), the Database of Abstracts of Reviews of Effects (DARE), the Cochrane Central Register of Controlled Trials (CENTRAL), NHS Economic Evaluation Database (NHS EED) and the Health Technology Assessment Database (HTA)) and EconLit were searched on 24th January 2018. The searches were restricted to publications from 2016 onwards.

The database searches were supplemented by a manual search of the Health Economics Research Centre Database of Mapping Studies from 2016 onwards. In addition, the company searched the reference lists of any identified systematic reviews.

The searches were designed to update previous quality of life searches for PsA carried out for TA445 in February 2016. The date limit restriction applied to the searches reported in the submission is appropriate to identify any new studies regarding HRQL in PsA published during the period 2016 to 2018. The searches were fit for purpose, conducted correctly and are clearly reported.

5.1.2 Inclusion/exclusion criteria used for study selection

Population

Inclusion criteria: Adult patients with active PsA who have had an inadequate response or who have been intolerant to a prior disease-modifying anti-rheumatic drug (DMARD) therapy

Exclusion criteria: Patients suffering from other rheumatic conditions.

Interventions/Comparators

Inclusion criteria: Tofacitinib, Biologic DMARDs (abatacept SC injection/IV infusion, adalimumab SC injection, etanercept SC injection, golimumab SC injection, infliximab IV infusion, certolizumab pegol SC injection, ustekinumab SC injection, secukinumab SC injection, ixekizumab SC injection)

and PDE-4 inhibitor (apremilast administered orally).

Exclusion criteria: Diagnostics. No restrictions placed on dosing regimen, including whether the treatments are used as monotherapy or in combination with another treatment or whether the UK-licensed dose is used.

Outcomes

Inclusion criteria: cost in combination with any of the following; LYGs, QALYs, DALYs.

No exclusion criteria specified for this domain.

Study design

Inclusion criteria: Comparative economic evaluations including cost-effectiveness analyses, cost-utility analyses, cost-benefit analyses, cost-minimisation analyses, cost-minimisation analyses, cost-consequence studies and economic evaluations of a single cohort.

Exclusion criteria: case reports and case studies. Editorials and any other non-systematic reviews.

Publication type

Systematic reviews of economic evaluations were included at the title/abstract screening stage and used for identification of any additional primary studies not identified through the database searches but were excluded during the full-text review stage.

Language restrictions

Inclusion criteria: English

Exclusion criteria: Non-English

The inclusion and exclusion criteria for the systematic review were supported by the rationale for each criteria as provided in Table G5 in Appendix G. However, excluding non-English language papers

means relevant foreign language papers may have been missed.

5.1.3 Studies included and excluded in the cost effectiveness review

No previously published cost-effectiveness studies of tofacitinib for PsA were identified.

The systematic review identified 17 evaluations that met the inclusion criteria. Fourteen of these were UK publications and the remaining 3 were non-UK evaluations which were deemed not relevant for decision-making in England. Of the 14 UK publications, 3 were NICE HTA monographs, 2 were NICE ERG reports, 3 were UK HTA review articles and 6 were some other form of UK evaluation.

Table 25 describes the UK publications that met the inclusion criteria and identifies the type of publication.

Table 25 Studies included in the cost-effectiveness review

Year	Author	Title	Type of publication
2006	Bansback et al	Estimating the cost and health status consequences of treatment with TNF antagonists in patients with psoriatic arthritis.	Other form of UK evaluation
2006	Woolacott et al	Etanercept and infliximab for the treatment of psoriatic arthritis: a systematic review and economic evaluation.	HTA monograph for TA104
2007	Bravo Vergel et al	The cost-effectiveness of etanercept and infliximab for the treatment of patients with psoriatic arthritis.	Other form of UK evaluation
2011	Cummins et al	Cost-effectiveness of infliximab for the treatment of active and progressive psoriatic arthritis.	Other form of UK evaluation
2011	Rodgers et al	Etanercept, infliximab and adalimumab for the treatment of psoriatic arthritis: a systematic review and economic evaluation.	HTA monograph for TA199

2011	Bojke et al	Modelling the cost-effectiveness of biologic treatments for psoriatic arthritis.	Other form of UK evaluation
2012	Yang et al	Golimumab for the Treatment of Psoriatic Arthritis: A NICE Single Technology Appraisal.	Review article
2012	Cummins et al	Cost effectiveness of golimumab for the treatment of active psoriatic arthritis.	Other form of UK evaluation
2014	Cawson et al	Systematic review, network meta-analysis and economic evaluation of biological therapy for the management of active psoriatic arthritis.	Other form of UK evaluation
2016	O'Connor et al	The Clinical and Cost Effectiveness of Ustekinumab for the Treatment of Psoriatic Arthritis: A Critique of the Evidence.	Review article
2016	Sideris et al	The Clinical and Cost Effectiveness of Apremilast for the Treatment of Psoriatic Arthritis: A Critique of the Evidence.	Review article
2017	Corbett et al	Certolizumab pegol and secukinumab for treating active psoriatic arthritis following inadequate response to disease-modifying antirheumatic drugs: a systematic review and economic evaluation.	HTA monograph for TA445

As described in Section 5.1.1, the company performed a search of the HTA websites. This search revealed the following appraisals: NICE (n=5 complete; n=1withdrawn; n=3 in progress), SMC (n=8), PBAC (n=6) and CADTH (n=6) ranging from 2005 to 2017. Tables G28-G33 in Appendix G summarise each of the identified TAs but it is unclear from the CS or Appendix how the company incorporated the result of this HTA search into their review.

5.1.4 Conclusions of the cost effectiveness review

Aside from the exclusion of non-English language papers, the search strategies were well specified and the searches appear to have been conducted appropriately.

The review identified a number of previous economic models but as mentioned in Section 5.1.3, no previous models were found which included tofacitinib as a comparator. Most of the evaluations identified were developed for, or based on those developed for, NICE technology appraisals. [1, 14-16]The company performed a quality assessment of the included studies and provided this in Appendix G (Tables G19-G27). The majority of the models adopted the same structure, and the company chose a similar structure to model the cost-effectiveness of tofacitinib.

It is clear from the systematic review that TA445 is the most comparable economic evaluation to the company's submission. However, the company does not explicitly identify this in the CS.

5.2 ERG's summary and critique of company's submitted economic evaluation

An overall summary of the company's approach and references to the relevant sections in the CS are reported in Table 26 below.

Table 26 Summary of the Company's economic evaluation (and signposts to company's submission)

Element	Approach	Source/Justification	CS reference
Model	A Markov model with 40 year time horizon and a 3-month cycle length. The model evaluates the cost-effectiveness of tofacitinib versus NICE-recommended comparators. The model reflects initial response to treatments, continued use or withdrawal from the treatment. Both the skin and joint symptoms of PsA are taken into account.	The model structure, methods and assumptions are reflective of current NICE guidance.	Section B3; p115
States and events	Response to treatment was evaluated according to PsARC response three months from baseline for all comparators. Non-responders transitioned to the subsequent treatment in the pathway; responders were assumed to continue treatment until they withdrew due to either a loss of efficacy, adverse events or death. Transitions from the treatment state to alternative pathways were determined by initial response rates and discontinuation rates. Adverse events were not modelled.	Transition response criteria according to BSR guidance and company's NMA. Withdrawals based on recent NICE guidance and [1].	Section B3; p115
Population and subgroups	Adults with active PsA whose disease has not responded adequately to previous DMARD therapy or for whom DMARDs are not tolerated or contraindicated.		Section B.1.1, Table 1, p;12 Section B.3.2.1. p113-114.

	The baseline characteristics were sourced from the tofacitinib trial population which includes both patients who had previously received biologic therapies and biologic-naïve patients. Four sub-populations were defined: 1) People whose disease has not responded adequately to 1 non-biological DMARD (Results not submitted for sub-population 1) 2) People whose disease has not responded adequately to at least 2 non-biological DMARDs 3) People whose disease has not responded adequately to non-biological DMARDs and 1 or more TNFis 4) People in whom TNFi are contraindicated or not tolerated.	The drug company seek to align the sup-populations assessed in the TA of tofacitinib for treating active PsA with cDMARDS to the populations that have received positive recommendations from NICE in previous TAs (i.e. sup-populations 2, 3 and 4)	
Comparators	Sequences of treatments are modelled. These include the comparator technologies: TNFis (adalimumab, certolizumab pegol, etanercept, golimumab and infliximab), secukinumab (an IL 17A inhibitor), ustekinumab (an IL12/IL23 inhibitor), apremilast (a PDE4 inhibitor), and best supportive care (BSC).	The NICE scope lists certolizumab pegol as a comparator for sub-population 3, which includes people whose disease has not responded adequately to non-biological DMARDs and 1 or more TNFi. Certolizumab pegol has been excluded from sub-population 3 because the data available from the RAPID PsA trial informs only a subset of patients in this sub-population (i.e., primary responders to a prior TNFi who were secondary failures [primary non-responders were explicitly excluded from this trial]) [1].	Section B.3.2.3; p118-119
Natural history	For patients receiving BSC or csDMARDS, a HAQ progression rate of 0.077 per year was applied. Patients can reach a maximum score of 3.	Obtained from NICE PsA guidance[14, 15] as estimated from Norfolk Arthritis Register [17].	Section B.3.3.1.5; P125
Treatment effectiveness	Criterion for continuing treatment was the probability of PsARC response, assessed at 12 weeks. Following initial response (or non-response) to treatment at 12 weeks, the arthritis and psoriasis-specific components of PsA are modelled separately.	Obtained from the company's NMA.	Section B.3.2.2; p114-117

	The arthritis component was modelled via a change in HAQ-DI score conditional on PsARC at 12 weeks. The psoriasis component was modelled via changes in PASI score at 12 weeks.		
Effectiveness of subsequent lines of therapy	For comparisons involving more than one line of treatment, subsequent treatments are assumed to be as efficacious as first line, i.e. no effect degradation is assumed.		
Discontinuation	12-week probability of withdrawal of 3.96% was included in the model. Patients who discontinued a treatment transitioned to the next treatment option or BSC when they had failed all treatments. Rebound to the baseline HAQ value was assumed for patients entering BSC (termed as rebound to initial gain).	NICE PsA Guidance as obtained from the York model [14].	Section B.3.3.1.; p 125
Adverse events	Adverse events were not explicitly modelled. AEs were only considered implicitly in terms of their effect on initial response and withdrawal for each treatment.	NICE PsA guidance as obtained from Corbett et al in 2017 [1].	
Mortality	Mortality rates were derived from life table for England and Wales (2014-2016).	A standardised mortality rate (1.36) reported by Ali et al [18] and as applied in TA445 was used [15].	Section B.3.3.1.7 p; 129
Health-related quality of life	Patients HRQoL is defined in the model in terms of HAQ and PASI scores, and these are mapped to EQ-5D. Patients HAQ-DI and PASI scores change according to treatment response. HAQ-DI scores remain constant while patients are on treatment with bDMARDS or tofacitinib but progress linearly while patients are on apremilast or BSC (reflecting worsening of physical functions following failure to respond to treatment. PASI scores do not progress on BSC as they are not progressive. Whilst on treatment, improvements in PASI scores are possible.	In the base case analysis utilities were based on a linear regression. A utility model based on tofacitinib trial data was used in scenario analysis and applied to either tofacitinib alone, or to tofacitinib and its comparators.	Section 3.4.2. p; 130-131
Resource utilization and costs	Costs included were: drug acquisition costs; drug administration costs and monitoring costs. Arthritis and psoriasis-related costs were also applied in the model and based on the HAQ-DI and PASI scores.	Resource use associated with drug administration and monitoring costs were obtained from the BNF [19] and TA199 and TA445, respectively [14, 15]. Acquisition costs were taken from the BNF and electronic market information tool (eMIT)	Section B.3.5. p;133-142

	Costs for the following treatments differ between the first cycle and subsequent cycles to account for loading doses or PAS arrangements; Apremilast, Certolizumab Pegol, Infliximab, Secukinumab and Ustekinumab.	database [19, 20]. No drug costs are assumed for BSC. Patient Access Scheme prices are listed where information is in the public domain. Administration and monitoring costs (except for liver function text, chest x-ray and TB heaf test costs)* were obtained from the NHS reference costs and PSSRU [21, 22]. Arthritis-related costs were estimated as a function of HAQ-DI score, based on Rodgers et al. Psoriasis-related costs based on PASI scores were obtained from TA445 [15]. Adverse events were not considered in the model	
Discount rates	3.5% for utilities and costs	NICE reference case	Section B.3.2.2. p; 117
Sensitivity analysis	Probabilistic sensitivity analysis and scenario analysis were performed. Deterministic sensitivity analysis were not performed.	Deterministic sensitivity analysis was not performed.	Section B.3.8.1. p;147 Section B.3.8.2 p;154

*Obtained from TA445

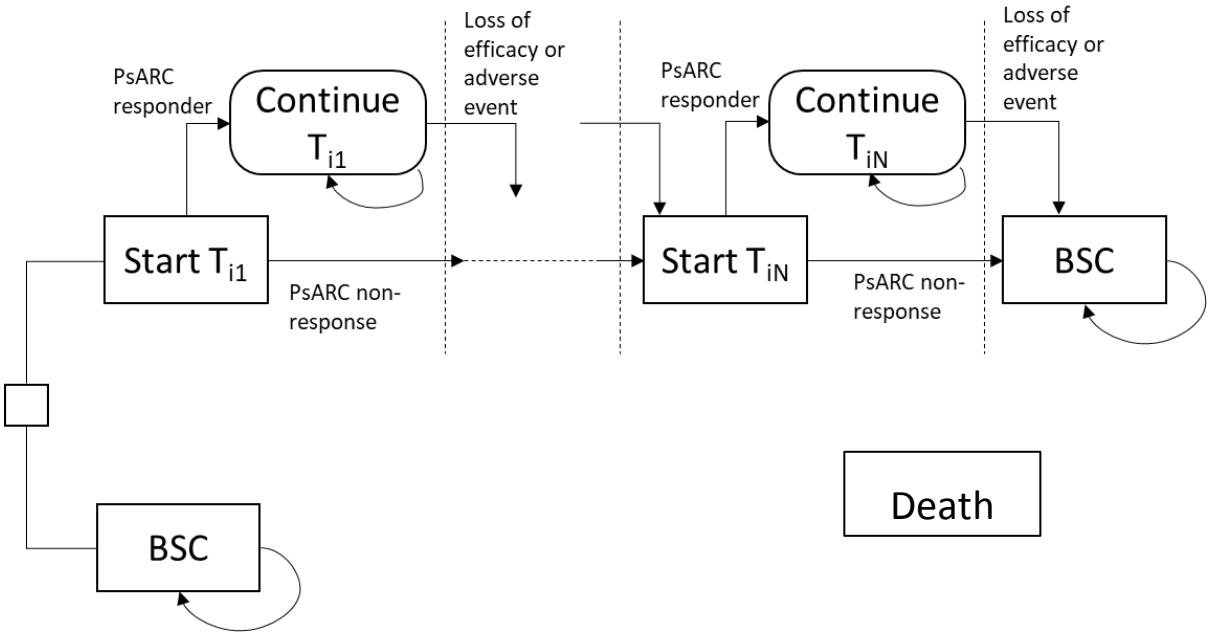
5.2.1 Model structure

The company describes a de novo economic evaluation based on a Markov cohort model similar to the model structure used by the York Assessment Group (AG) in TA445 [15]. The model was developed in Microsoft Excel to evaluate the cost-effectiveness of tofacitinib. The model structure allows a comparison of multiple treatment sequences (see Section 5.2.4). The model allows patients to cycle through sequences of therapy, with patients remaining on a treatment after the first 3 months if they have met the required criteria.

After an initial response to treatment, patients remain on therapy until either a loss of efficacy, the occurrence of particular adverse events or death. Transition to death (all cause and excess due to PsA) is included at each cycle of the model.

A schematic representation of the company's model is shown in Figure 7. Rather than specifying health states, between which patients transition, the company defines states relating to which treatment is being received and if this is during the primary response or maintenance phase.

Figure 7 Model Summary (Figure 15, p115 in CS)



Patients may transition to the death state from any other state. Abbreviations: BSC, best supportive care; PsARC, Psoriatic Arthritis Response Criteria; T_{I1}, first therapy in the i^{th} sequence; T_{iN}, n^{th} therapy in the i^{th} sequence.

In the base case model, Psoriatic Arthritis Response Criteria [PsARC] response at 3 months is used to determine the proportion of patients remaining on treatment. This reflects the clinical management of PsA as recommended by NICE [14, 15, 23, 24] and the BSR [25]. A PsARC response is binary, representing the proportion of people who respond or do not respond to treatment. The psoriasis component of PsA is modelled via changes in Psoriasis Area and Severity Index (PASI) scores, these are defined as the proportion of patients with a 50, 75 and 90% change in their baseline PASI score. In the base case model, it is assumed that PASI change does not determine treatment continuation, thus only PsARC scores are used as the response criteria. PASI response is assumed to be correlated with PsARC responses (Section B3.3.2.1 in CS). Conditional on PsARC response, patients were categorised as PASI-75 (See Section 5.2.3) responders and non-responders, respectively.

Following the initial response (or non-response) to treatment at 3 months, the psoriasis- and arthritis-specific components of PsA are modelled separately. The arthritis component of PsA is modelled via a change in the Health Assessment Questionnaire-Disability Index (HAQ-DI) score conditional on PsARC response at 3 months. Mean changes in HAQ-DI scores for PsARC responders and non-responders were treatment specific and taken from the NMA (Section B.3.2.2 in CS). For PsARC responders, HAQ-DI change from baseline is maintained beyond 3 months in line with previous

modelling approaches, such as that adopted by the AG in TA445 [15] (Section B.3.2.2 in CS), with the exception of apremilast (as per TA433 [26]) and best supportive care (BSC), whereby HAQ scores increase in a linear fashion (see Figure 8 and Section B.3.2.2 in CS).

Figure 8 HAQ score changes over time (Figure 16, p116 in CS)

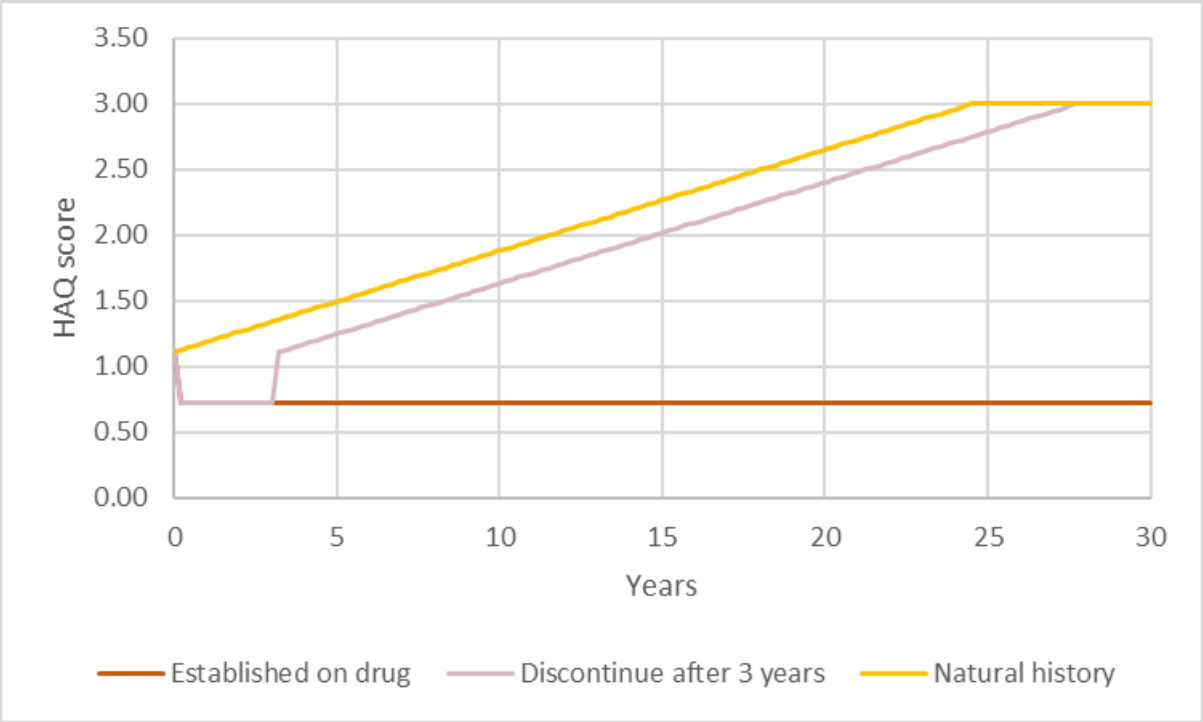


Figure 8 illustrates the progression HAQ-DI over time for three types of patients: a patient successfully established on a bDMARD; a patient discontinuing after 3 years (and transitioning to BSC); and a patient receiving BSC. When patients discontinue treatment, it is assumed that they experience a rebound in HAQ-DI and PASI scores equal to their initial gain. These assumptions are in line with the York AG model [1] from TA445 [15].

For those remaining on treatment (responders) an assumption of no HAQ-DI progression was made for the 'continued use' health state (see Section 5.2.6.3). Patients who discontinued a treatment and transitioned to the next treatment option were assigned the HAQ-DI score for PsARC non-responders receiving the previous treatment for the duration of the trial period of the current treatment, after which they rebound to their starting HAQ-DI score. The psoriasis component of PsA is assumed to be non-progressive and therefore PASI scores do not increase while patients remain on therapy or BSC.

For those patients that progress to BSC the HAQ-DI rebounds back to the pre-treatment level (see Figure 5.1), which is consistent with the rebound equal to gain applied in previous economic models.

In addition HAQ-DI subsequently increases at a rate consistent with the natural history of PsA in patients who receive no treatment up to a maximum value of 3. This assumption has been applied in previous economic models in PsA [14]. It is not made explicit in the CS what happens to HAQ-DI post 3 months for non-responders, however in the electronic model this appears to equate to a rebound back to starting HAQ-DI.

Similar to previous models (TA445), a scenario is specified where disease activity is modelled using the American College of Rheumatology response criteria (ACR20/50/90). For example, an ACR20 response is defined as a 20% reduction in ACR, with corresponding terminology used for alternative percentage reductions (e.g., ACR 50 and ACR 70 for 50% and 70% reductions in ACR, respectively) (Section B.3.3.1.4.). The company model allowed additional alternative response scenarios: PASI alone and PASI and PsARC response. The results of these scenarios are not presented in the CS.

5.2.2 The company's economic evaluation compared with the NICE reference case checklist

Table 27 summarises the economic submission and the ERG's assessment of whether the *de novo* evaluation meets NICE's reference case.

Table 27 NICE reference case list

Attribute	Reference Case	Included in CS	Comment on whether de novo evaluation meets requirement of NICE reference case
Comparator(s)	As listed in the scope developed by NICE	Partly	Omitted sub-population 1 (People whose disease has not responded adequately to 1 non-biological DMARD
Perspective on outcomes	All direct health effects, whether for patients or, when relevant, carers	Yes	QALY benefits to patients treated were considered
Perspective on costs	NHS and PSS	Yes	NHS and PSS costs were taken into account
Type of economic evaluation	Cost-effectiveness analysis with fully incremental analysis	Yes	A Markov cohort model was employed for the cost-effectiveness analysis. The model compared the costs and QALY outcomes of treatment sequences.
Time horizon	Long enough to reflect all important differences in costs or outcomes between the technologies being compared.	Yes	A 40 year time horizon was adopted, consistent with recent published cost-effectiveness analyses. PsA is a chronic, lifetime condition with no known cure. Disease management aims to improve symptoms and HRQoL over a patients' lifetime. A 40 year time horizon accounts for

			the long-term consequences of the disease. However, long-term time-horizons rely on assumptions, due to the lack of long-term data.
Synthesis of evidence on health effects	Based on systematic review	Yes	In the absence of head-to-head trials between the identified comparators, a network meta-analysis was conducted to inform the clinical efficacy parameters in the economic model
Measuring and valuing health effects	Health effects should be expressed in QALYs. The EQ-5D is the preferred measure of health-related quality of life in adults.	Yes	A regression equation was used which maps HAQ-DI and PASI scores to EQ-5D. The algorithm generated as part of TA445 [15] was used. Regression coefficients calculated using the EQ-5D results from the tofacitinib trial were only tested in sensitivity analysis and applied to all treatments. (CHECK THIS)
Source of data for measurement of health-related quality of life	Reported directly by patients and/or carers	Yes	
Source of preference data for valuation of changes in health-related quality of life	Representative sample of the UK population	Yes	Utility values were based on ED-5D estimates.
Equity considerations	An additional QALY has the same weight regardless of the other characteristics of the individuals receiving the health benefit	Yes	
Evidence on resource use and costs	Costs should relate to NHS and PSS resources and should be valued using the prices relevant to the NHS and PSS	Yes	
Discounting	The same annual rate for both costs and health effects.	Yes	Costs and benefits were discounted at 3.5%.

5.2.3 Population

The CS defined the target population for the base case analysis as patients with active PsA whose disease has not responded adequately to previous DMARD therapy or for whom DMARDs are not tolerated or contraindicated. Four sub-populations were outlined in the NICE scope:

1). People whose disease has not responded adequately to 1 non-biological DMARD.

2). People whose disease has not responded adequately to at least 2 non-biological DMARDs.

3). People whose disease has not responded adequately to non-biological DMARDs and 1 or more TNFis

4). People in whom TNFi are contraindicated or not tolerated.

The company sought to align the sup-populations assessed in this technology appraisal (TA) to the populations that have received positive recommendations from NICE in previous TAs (i.e. sup-populations 2, 3 and 4). As a result, the company did not submit results for sub-population 1.

The base case of the company's economic model included patient data from two key Phase III clinical trials, OPAL Broaden and OPAL Beyond (See Section 4). Sub-populations 2 and 4 were informed by the bDMARD-naïve evidence synthesis with data for tofacitinib from OPAL Broaden (csDMARD-IR and TNFi-naïve); and subpopulation 3 was informed by the bDMARD-experienced evidence synthesis with data for tofacitinib from OPAL Beyond (TNFi-IR). Patient characteristics in the tofacitinib trials are discussed in more detail in Section 5.2.6.1.

For all sub- populations (2, 3 and 4), baseline psoriasis is derived from data reported by the British Association of Dermatologists [27]. As per TA445, the population is split into 50% with no psoriasis, 25% with mild to moderate psoriasis, and 25% with moderate to severe psoriasis. In TA445, PASI response was assessed separately for each sub-group defined by its baseline level of psoriasis; no psoriasis (baseline PASI = 0.00), mild to moderate psoriasis (baseline PASI = 7.3) and moderate to severe psoriasis (baseline PASI = 12.5). In the company's model however, a weighted average PASI score of these three subgroups was calculated for the entire population, for each model cycle, therefore sub-populations were not defined according to psoriasis level. It is important to explore this assumption given the impact that differences in baseline characteristics such as HAQ-DI, and particularly PASI scores can have on cost-effectiveness results. More importantly, the severity of psoriasis determines the appropriate dosing of the comparator secukinumab; where secukinumab 300mg is approved for patients with severe psoriasis as opposed to the standard does of secukinumab 150mg. This assumption is explored in section 6.

5.2.4 Interventions and comparators

All technologies included in the cost-effectiveness analysis i.e. TNFis (adalimumab, certolizumab pegol, etanercept, golimumab and infliximab), IL inhibitors (secukinumab and ustekinumab) and PDE4 inhibitor (apremilast) were modelled in line with their marketing authorisation. BSC is also included as a comparator for each sub-population and is representative of the placebo arm of the clinical trials included in the NMA, therefore assuming the same efficacy of placebo. No separate

costs are associated with BSC as these costs are assumed to be captured in the estimates of resource use associated with HAQ-DI. It was unclear in the CS how BSC was defined, the ERG asked for clarification on this. In their response to clarification, the company defined BSC as a mixture of csDMARDs and/or usual care (e.g.NSAIDs, corticosteroids). They state that BSC reflects the clinical effectiveness estimates of the placebo groups in the trials of tofacitinib and the relevant comparators of the NMA. They justify their definition as being consistent with TA445 [15].

For tofacitinib, a dosage of 5mg twice daily was assumed, taken orally. The included comparators and their respective dosage regimens are listed below:

- adalimumab 40mg given every other week, administered as a subcutaneous injection
- certolizumab pegol 200mg every other week, administered as a subcutaneous injection
- etanercept 25mg twice weekly, administered as a subcutaneous injection
- golimumab 50mg once a month, administered as a subcutaneous injection
- inflixumab 5mg/kg of body weight every 8 weeks, administered as an intravenous infusion
- secukinumab 150mg once a month, administered as a subcutaneous injection
- secukinumab 300mg once a month, administered as a subcutaneous injection
- secukinumab weighted dose once a month, administered as a subcutaneous injection
- ustekinumab 45mg every 12 weeks, administered as a subcutaneous injection
- apremilast 30mg twice daily, taken orally

The selection of the first treatment in a sequence for each sub-population is based on previous NICE recommendations [14, 15, 23, 24, 26] and the NICE scope. The selection of the second and third treatment options reflects TA445 [15]. As some sub-populations are eligible for more lines of treatment (prior to moving to BSC) than others, the length of treatment sequence varies across the sub-populations.

In terms of the comparators, the final scope issued by NICE included different comparators for different patient populations (see Table 28).

Table 28 Treatment sequences for each patient sub-population (Table 42, p119 in CS)

Patient sub-population	Treatment option as per NICE scope†		
	First in sequence	Second in sequence	Third in sequence
Sub-population 2: Disease has not responded to at least 2 nbDMARDs*	TOF	UST	BSC
	ADA		
	APR		
	CZP		
	ETN		

	GOL		
	INF		
	SEC (188mg, weighted dose)		
	BSC	-	-
Sub-population 3: Disease has not responded to nbDMARDs and at least 1 TNFi	TOF	BSC	-
	SEC (300mg)		
	UST		
	BSC	-	-
Sub-population 4: TNFi contraindicated or not tolerated	TOF	BSC	-
	SEC (188mg, weighted dose)		
	UST		
	BSC	-	-

†First treatment in sequence options are chosen in accordance with NICE guidance [14, 15, 23, 24, 26]. Second- and third treatment in sequence options are aligned with those used in TA445[15].*nbDMARDs ~ csDMARDs

Abbreviations: ADA, adalimumab; APR, apremilast; BSC, best supportive care; bDMARD, biological disease-modifying anti-rheumatic drug; CZP, certolizumab pegol; ETN, etanercept; GOL, golimumab; INF, infliximab; nbDMARD, non-biological disease-modifying anti-rheumatic drug; SEC, secukinumab; TNFi, TNF inhibitor; TOF, tofacitinib; UST, ustekinumab.

The NICE scope lists CZP as a comparator for sub-population 3, which includes people whose disease has not responded adequately to non-biological DMARDs and 1 or more TNFi. Similar to TA445, the company excluded CZP from sub-population 3 because the data available from the RAPID PsA trial informs only a subset of patients in this sub-population (i.e., primary responders to a prior TNFi who were secondary failures [primary non-responders were explicitly excluded from this trial]) [1].

For all sub-populations, following a lack of response to PsARC or subsequent withdrawal for PsARC responders, patients moving onto the next line of treatment, are assumed to have the same response probabilities as first line treatment, i.e. no effect degradation is applied for subsequent lines of therapy. The ERG has concerns that the CS does not address the issue of effect degradation for subsequent lines of treatment in the model and question the validity of this assumption [15] . This assumption is discussed further in Section 5.2.6.2.

5.2.5 Perspective, time horizon and discounting

The perspective of the company's de novo economic analysis was the NHS and Personal Social Services and an annual discount rate of 3.5% for both costs and health effects was applied, in line with the NICE reference case.

The time horizon of the model was 40 years, which was stated to be consistent with the most recent published cost-effectiveness analyses in PsA and accounts for the long-term consequences of a chronic, lifetime disease like PsA.

A 3 month cycle length was used. Given the short cycle length, a half-cycle correction was not applied.

5.2.6 Treatment effectiveness and extrapolation

The details of the effectiveness data used in the economic model are discussed in Sections 5.2.6.1. to 5.2.6.6.

5.2.6.1 Baseline Patients Characteristics

The baseline patient characteristics applied in the model were sourced from the tofacitinib clinical trials. Given that the company implemented a network meta-analysis (NMA) to inform treatment efficacy parameters for all treatment in the economic model, the ERG considers that the set of studies included in the NMA could be a more appropriate evidence base to inform the baseline characteristic of the patient population.

Baseline patient characteristics from the NMA are not included in the CS or Appendices for comparison. Instead, Table 29 provides a comparison of baseline characteristics between the bDMARD-naïve and bDMARD-experienced population from the OPAL Broaden and Beyond tofacitinib trial with the baseline characteristics used previously in TA445.

Table 29 Comparison of baseline characteristics

Description	CS (bDMARD-naïve)	CS (bDMARD-experienced)	TA445
Baseline age	47.9	50	47
Gender (% female)	53%	55%	Not identified
Baseline HAQ	1.11	1.30	1.22
Baseline PASI for no psoriasis	0.0	0.0	0.0
Baseline PASI for mild to moderate psoriasis	7.3	7.3	7.3
Baseline PASI for moderate to severe psoriasis	12.5	12.5	12.5

The ERG requested justification for the use of baseline characteristics from the tofacitinib trials (as opposed to the baseline patient characteristics from the trials included in the NMA). The ERG also requested a scenario using the baseline patient characteristics from the NMA.

The company justified the use of baseline characteristics from the trials as the most representative of the populations under consideration due to changes in standard of care since some of the previous trials were conducted, particularly in terms of prior treatments. For example, the company identify that the trials conducted by Mease et al (2000) and the ADEPT trial in 2005 do not include inadequate response to previous DMARD in their inclusion criteria. This does form part of the inclusion criteria in the OPAL trials therefore reflecting the changes in standard of care over time. The company provided the results of the scenario analyses using the baseline patient characteristics from the NMA. Tables 14 to 16 in the company's response to clarification provide these results, which results in only a small change in the ICERs for subpopulations 2-4.

5.2.6.2 Response Rates

In the absence of head-to-head trial data, response rates for all treatments included in the model were obtained primarily from the company NMA. Three outcomes were included in the NMA to inform the economic model; (1) PsARC response, (2) change in HAQ-DI score conditional on PsARC response and (3) PASI 50, PASI 75 and PASI 90 responses. The probability of PsARC response for the bDMARD-naïve and bDMARD-experienced population as implemented in the model and changes in HAQ-DI score conditional on PsARC response are presented in Table 30. Probabilities of PASI 50, PASI 75 and PASI 90 responses for the bDMARD-naïve and bDMARD-experienced population as implemented in the model are reported in Table 31.

In the base case, the model uses PsARC response rates at 3 months to determine the proportion of patients remaining on treatment. This reflects the clinical management of PsA as recommended by NICE [14, 15, 23, 24, 26, 28]. The 3 month cycle length is also reflective of the continuation rule which means that patients must achieve a PsARC response within 3 months to remain on therapy. This continuation rule is in line with guidance from the BSR [29] and previous NICE appraisals [14, 15, 23, 24, 26]. However, this does not reflect the continuation rule for all comparators in the model e.g. APR and SEC, according to their SPCs, should be assessed at week 16 and UST at 24 weeks.

For the bDMARD-experienced population (sub-population 3), in the NMA, it was only possible to estimate PsARC response for tofacitinib 5 mg BD, ustekinumab and placebo due to a lack of response data available in primary and secondary publications. To include PsARC response for secukinumab in the economic model, the odds ratio for secukinumab 300 mg versus placebo was taken from the base-case analysis for the bDMARD-experienced population from TA445 [15]. HAQ-DI change conditional on PsARC response was not available in either the naïve or experienced populations for secukinumab and certolizumab, therefore the values from the TA445 meta-regression NMA of HAQ scores have been incorporated into the model for these comparators in the bDMARD naïve populations. In the

bDMARD experienced population the values have been taken from the TA445 [15] and bDMARD experienced NMA.

Consistent with previous economic models (TA199 [14] and TA445 [15]), it was assumed that PASI-75 response rates may vary by treatment response (based on PsARC). In order to capture this, a positive correlation between PsARC and PASI-75 response was included in the model. The company adopted the correlation coefficient between PsARC and PASI-75 (0.436), as used in the York model in TA199 [14] and TA445 [15].

Table 30 Summary of PsARC response probabilities and HAQ-DI absolute score changes

Variable	bDMARD-naïve population	bDMARD-experienced population
Probability of PsARC Response		
Placebo	[redacted]	[redacted]
Adalimumab	[redacted]	[redacted]
Apremilast	[redacted]	[redacted]
Etanercept	[redacted]	[redacted]
Infliximab	[redacted]	[redacted]
Ustekinumab	[redacted]	[redacted]
Golimumab	[redacted]	[redacted]
Tofacitinib 5mg	[redacted]	[redacted]
Secukinumab 150mg	[redacted]	[redacted]
Secukinumab 300mg	[redacted]	[redacted]
Certolizumab	[redacted]	[redacted]
Ixekizumab 80 Q2W	[redacted]	[redacted]
Ixekizumab 80 Q4W	[redacted]	[redacted]
HAQ-DI score change for PsARC responder		
Placebo	[redacted]	[redacted]
Adalimumab	[redacted]	[redacted]
Apremilast	[redacted]	[redacted]
Etanercept	[redacted]	[redacted]
Infliximab	[redacted]	[redacted]
Ustekinumab	[redacted]	[redacted]
Golimumab	[redacted]	[redacted]
Tofacitinib 5mg	[redacted]	[redacted]
Secukinumab 150mg	[redacted]	[redacted]
Secukinumab 300mg	[redacted]	[redacted]
Certolizumab	[redacted]	[redacted]
Ixekizumab 80 Q2W	[redacted]	[redacted]
Ixekizumab 80 Q4W	[redacted]	[redacted]
HAQ-DI score change for PsARC non-responder-		
Placebo	[redacted]	[redacted]
Adalimumab	[redacted]	[redacted]
Apremilast	[redacted]	[redacted]
Etanercept	[redacted]	[redacted]
Infliximab	[redacted]	[redacted]
Ustekinumab	[redacted]	[redacted]

Golimumab	[redacted]	[redacted]
Tofacitinib 5mg	[redacted]	[redacted]
Secukinumab 150mg	[redacted]	[redacted]
Secukinumab 300mg	[redacted]	[redacted]
Certolizumab	[redacted]	[redacted]
Ixekizumab 80 Q2W	[redacted]	[redacted]
Ixekizumab 80 Q4W	[redacted]	[redacted]

Table 31 Summary of PASI-50, PASI-75 and PASI-90 response probabilities

Variable	bDMARD-naïve population	bDMARD-experienced population
Probability of PASI-50 response		
Placebo	[redacted]	[redacted]
Adalimumab	[redacted]	[redacted]
Apremilast	[redacted]	[redacted]
Etanercept	[redacted]	[redacted]
Infliximab	[redacted]	[redacted]
Ustekinumab	[redacted]	[redacted]
Golimumab	[redacted]	[redacted]
Tofacitinib 5mg	[redacted]	[redacted]
Secukinumab 150mg	[redacted]	[redacted]
Secukinumab 300mg	[redacted]	[redacted]
Certolizumab	[redacted]	[redacted]
Ixekizumab 80 Q2W	[redacted]	[redacted]
Ixekizumab 80 Q4W	[redacted]	[redacted]
Probability of PASI-75 response		
Placebo	[redacted]	[redacted]
Adalimumab	[redacted]	[redacted]
Apremilast	[redacted]	[redacted]
Etanercept	[redacted]	[redacted]
Infliximab	[redacted]	[redacted]
Ustekinumab	[redacted]	[redacted]
Golimumab	[redacted]	[redacted]
Tofacitinib 5mg	[redacted]	[redacted]
Secukinumab 150mg	[redacted]	[redacted]
Secukinumab 300mg	[redacted]	[redacted]
Certolizumab	[redacted]	[redacted]
Ixekizumab 80 Q2W	[redacted]	[redacted]
Ixekizumab 80 Q4W	[redacted]	[redacted]
Probability of PASI-90 response		
Placebo	[redacted]	[redacted]
Adalimumab	[redacted]	[redacted]
Apremilast	[redacted]	[redacted]
Etanercept	[redacted]	[redacted]
Infliximab	[redacted]	[redacted]
Ustekinumab	[redacted]	[redacted]
Golimumab	[redacted]	[redacted]
Tofacitinib 5mg	[redacted]	[redacted]
Secukinumab 150mg	[redacted]	[redacted]
Secukinumab 300mg	[redacted]	[redacted]
Certolizumab	[redacted]	[redacted]
Ixekizumab 80 Q2W	[redacted]	[redacted]

Ixekizumab 80 Q4W	[redacted]	[redacted]

The ERG identified discrepancies in several of the efficacy results reported in the results of the NMA compared to the response rates used in the economic model. The results reported in the CS on the probability of PsARC, PASI and HAQ-DI conditional on PsARC response in the bDMARD-naïve population with ustekinumab do not match the NMA results that are used in the economic model. The majority of the results for PsARC, PASI and HAQ-DI conditional on PsARC response with all comparators reported in the CS for the bDMARD-experienced population do not correspond to those reported in the model. In terms of the ACR response rates, the company submission states that model E1 FE with 24-week data was selected as the 'pessimistic' case for the ACR endpoint and model E1 FE without 24-week data was selected as both the 'optimistic' and base case for the ACR response. However, in the model for all sub-populations, model E1 FE without 24-week data was selected for the base case and 'pessimistic' case while model E1 FE with 24-week data was used for the 'optimistic' case data. Additional details and justification were requested from the company.

The company reported that the differences in the NMA data between the economic model and company submission were caused primarily by differences in the outcomes reported. In the CS, the median values were presented (as per TA445), while the economic model uses the mean values. The company indicated that these are more appropriate for economic modelling (as per TA445). For the bDMARD naïve results, the company reported that the median values were mistakenly copied into the model in some instances (instead of the mean values from the NMA). The company identified that the values presented in the economic model for ustekinumab for the bDMARD naïve analysis were the mean values from the bDMARD experienced NMA (which was deemed the appropriate NMA as ustekinumab was second option in the treatment sequence, usually post bDMARD). The company stated that it was these factors that resulted in inconsistencies between data in the model and those in the main CS.

In their clarification response, the company reported that the economic model used the incorrect version of the ACR NMAs for the pessimistic and optimistic scenarios in the bDMARD-experienced population and that those reported in the submission are correct. As stated in Section 5.2.1, the ACR response (ACR20) was only used in scenario analyses and therefore do not affect the base case results.

In addition, when comparing the base case NMA models informing the effectiveness data included in the company's model, there were differences in the NMA models used in the current TA and the previous TA on which the current evaluation is based [15]. One reason for these variations is due to data that was previously publicly available for TA445 and no longer publicly available for the current TA.

More specifically, in terms of HAQ response, the company implemented a different base case model compared to that used in TA445. This differs from the TA445 [30] base case in that it uses random effects, adjusts for trials with more than two arms, and uses separate models for responders and non-responders. The analyses using separate models for responders and non-responders predict larger changes in HAQ-DI for responders than do the combined models, including for placebo responders. The ERG requested justification for this model specification. The ERG explores the validity of the NMA in Section 4.6 and explores the sensitivity of the economic model results to alternative NMA models in Section 6.2.

The response rates applied in the economic model assume that the treatment effect is maintained for subsequent lines of therapy, i.e. no reduction in effectiveness is applied for patients failing to respond to first line therapy or for those that initially respond but later withdraw due to loss of efficacy of adverse events. As discussed in TA445, this assumption is unlikely to be valid; however there is a paucity of data from which to estimate this effect degradation. For treatments with a lower PsARC response rate, i.e. higher number of patients moving onto 2nd line treatment, an assumption of no effect degradation may overestimate cost-effectiveness. Due to the lack of flexibility in the company model, the ERG is unable to explore the sensitivity of the cost-effectiveness results to this assumption in Section 6.

5.2.6.3 Natural history disease progression

As the psoriasis element of PsA is not progressive, the company assumes that PASI scores do not increase over time for patients receiving BSC. The arthritis element of PsA is assumed to be progressive, therefore, for patients not receiving biologic therapies (BSC), the company assumes the HAQ-DI score worsens overtime.

In the base case model the rate of progression for BSC was obtained from the York AG model [14]. This HAQ-DI progression was estimated based on an extract of data for PsA patients receiving palliative care included in the Norfolk Arthritis Register [17] until 2009. A worsening (increase) in HAQ-DI score of 0.077 per year was applied as the rate of natural disease progression in the company's economic model. Patients could reach a maximum HAQ-DI score of 3.

For biologic drugs, excluding apremilast, the company assumed no progression of disease whilst on treatment. The appraisal committee for TA433 concluded that there was insufficient evidence to demonstrate that apremilast halts radiographic disease progression (49), and concluded that the rate of disease progression experienced while receiving apremilast was assumed to be half of the progression rate for BSC/csDMARDs (i.e. 0.0385 per year). The same assumption was applied for apremilast in this analysis.

There is uncertainty about the trajectory of HAQ-DI over time, for both patients maintained on active therapies (responders) and those receiving BSC (either because of primary non-response or due to withdrawal).

Firstly, for patients receiving BSC they are assumed to follow a natural history trajectory through HAQ, with HAQ scores worsening at every cycle of the model. There are two main issues with this simplifying assumption. Firstly there appears to have been no attempts to update work from 2009 with a more recent extract from NOAR (or similar register such as ERAS). Practice regarding cDMARDs may change over time and this should be reflected in the HAQ change applied to the BSC comparator. In addition it is unlikely that the relationship between HAQ and time is linear over the entire extrapolation period (40 years). Recent work by Norton et al [31] looks at the progression of HAQ scores over 15 years in a largely RA population (but including some PsA patients in one dataset). This showed that HAQ progression becomes less linear over time, particularly post 5 years where scores stabilise.

For patients maintained on active therapies (responders), the CS assumes that patients responding to treatment do not progress further in terms of HAQ (full disease modification). The ERG has concerns regarding the validity of this assumption. As discussed in Section B.3.3.1 to assess the radiographic progression of tofacitinib 5mg BD, the company performed a population-adjusted analysis using pre-specified effect modifiers and prognostic factors centred using the baseline characteristics from the ADEPT trial, to adjust the OPAL Broaden data to a target population more at risk of progression. On the basis of this analysis, the company conclude that there are no differences between tofacitinib 5mg BD and adalimumab with respect to radiographic progression but that the analysis is limited given that the OPAL Broaden trial was underpowered to detect differences between tofacitinib 5mg BD and adalimumab. In addition, the company acknowledges that the prognostic factors for radiographic progression in the OPAL Broaden clinical trial were different (lower) (e.g., baseline CRP levels, baseline mTSS, baseline erosion and joint space narrowing scores) than a number of previous bDMARD studies in PsA [32]. As the evidence presented on radiographic progression is based on short-term follow-up and 11.3% of patients experience a progression (increase in mTSS) [33] the ERG considers that the rate of progression for tofacitinib is uncertain and therefore the ERG explores this assumption in Section 6.3.

5.2.6.4 Discontinuation

For PsARC responders, there is a risk of withdrawal following the first 3 months of treatment. Based on previous appraisals [14, 15], this probability is estimated from a meta-analysis of registry data from several countries to be -1.823. Withdrawal rates were assumed to be independent of HAQ-DI and PASI scores in the model. The same withdrawal rate is applied to tofacitinib and all comparators and

is assumed to be constant over time. The assumption of equal withdrawal rate is subject to uncertainty. This uncertainty is based on the mode of administration of tofacitinib and its impact on patient adherence. Following a discussion with its clinical advisor, the ERG are concerned regarding patient compliance with tofacitinib. As tofacitinib is an oral treatment taken twice daily, there is a possibility that patients may not take the drug appropriately and consistently over time.

In the CS, the company did not provide additional evidence or justification to support this assumption given the different mechanism and mode of delivery of tofacitinib. The ERG requested this information to be provided. More specifically, the ERG requested the withdrawal data for patients whose disease initially responds to treatment and subsequently discontinued treatment due to loss of efficacy of adverse events. The ERG requested a revised version of the model that allows a separate withdrawal rate to be specified for tofacitinib. Finally, the ERG request additional scenarios which use the withdrawal based on the data from the OPAL trials for tofacitinib.

In response, the company provided tables detailing the discontinuation rates and reasons for discontinuation among those whose disease initially responded to treatment in the OPAL Broaden and OPAL Beyond trials. They also provided a revised version of the model which allows a separate withdrawal rate to be specified for tofacitinib. The company provided an additional scenario using the rate of withdrawal from the OPAL trials for tofacitinib. However, the rate of withdrawal they used in the scenario analyses includes data for PsARC responders, PsARC non-responders and patients in which PsARC response data was missing at 3 months. The ERG recalculated the rate of withdrawal for PsARC responders using the data provided in response to clarifications (Table 00099.2.2.2 in company response). This suggests that the rate of withdrawal is around 5.5% per year, and therefore the ERG is satisfied that the base case assumption of equivalent withdrawal to the other biologics is valid.

5.2.6.5 Mortality

Mortality was not measured as an outcome in the tofacitinib clinical trials, so the treatment-specific impact on mortality was not assessed. The base case analysis of the economic model included all-cause, age-dependent probabilities of death based on the general England and Wales population from the national life tables published by the Office for National Statistics (ONS) [34]. The excess mortality risk associated with PsA is modelled using a HR of 1.36. This ratio was obtained from a prospective study of patients with PsA [18] and was applied in TA445 [15]. The ERG considers this to be a valid assumption.

5.2.6.6 Adverse Events

The incidence of adverse events leading to discontinuation from treatment was captured in the clinical trials for tofacitinib. Adverse events (AEs) are not explicitly included in the model, neither as a utility decrement nor as additional cost for their treatment. In the model, AEs were considered in terms of their effect on initial response and on the long-term rates of withdrawal from the continued use for each treatment. The ERG considers this to be a valid assumption.

5.2.7 Health related quality of life

Patients' HRQoL is defined in the model in terms of HAQ and PASI scores and these are mapped to EQ-5D. The health states in the model are defined by the treatment received and response to treatment. Patients' HAQ-DI and PASI scores remain constant while patients are on treatment with bDMARDs or tofacitinib, but they progress linearly while patients are on apremilast or BSC (reflecting worsening of physical function following failure to response to treatment (See Section 5.2.6.6).

EQ-5D data were available from the OPAL Broaden and OPAL Beyond clinical trials for tofacitinib. The company states that to be consistent with previous appraisals (TA119 [14] and TA445 [15]), the mapping algorithm used in the York model for the base case is implemented here. For the base case, the following formula from the York model was used:

Equation 5.1 Mapping algorithm

$$EQ-5D = 0.897 - 0.289 * HAQ - 0.004 * PASI$$

Scenario analysis was performed in which the *de novo* mapping algorithms derived using individual patient data (IPD) from the OPAL Broaden and Beyond clinical data were applied to tofacitinib alone or tofacitinib and its comparators.

Statistical models were developed using data from the OPAL Broaden (sub-populations 2 and 4) and OPAL Beyond (sub-population 3) studies separately. Two models were estimated using each study:

- A 'main effect' model predicting EQ-5D in which HAQ and PASI scores were included as independent covariates.
- An 'interaction effect' model which augmented the 'main effect' model by including the interaction between HAQ and PASI scores as a covariate.

Both models pool all non-missing data at all time points from across all arms of the respective clinical trials. Models were implemented as mixed effects models to account for repeated measures within subjects. The CS refers to Appendix Q for the results of these models but Appendix Q was not provided. The ERG requested this and also requested that the specific covariates and regression function be provided.

In addition, the CS does not provide the EQ-5D data as collected in the OPAL trials. The ERG requested results of any EQ-5D assessments in OPAL Broaden, OPAL Beyond and OPAL Balance including sample sizes, missing data, follow up points, EQ-5D scores at baseline and follow up for each treatment and details and results of any statistical tests performed.

In response, the company provided Appendix Q. Appendix Q details the specific covariates included in the company's scenario analyses which used the *do novo* mapping algorithm applied initially to all treatments and then to the tofacitinib arm only. The ERG compared the covariates used in these scenario analyses with those used in the previous models and conclude that the covariates are very similar to those used in previous appraisals (TA119 [14] and TA445 [15]). The company clarified that a mixed effects regression function was used to account for repeated measures in the data. The company provided tables reporting the EQ-5D assessments in the OPAL Broaden, OPAL Beyond and OPAL Balance tofacitinib trials. These tables described the average EQ-5D utilities up to 12 months for tofacitinib, tofacitinib 10mg BD, adalimumab, placebo, placebo → tofacitinib and placebo → tofacitinib 10mg BD, the change from baseline in EQ-5D utilities, EQ-5D utilities by PsARC response and the change in EQ-5D utilities from baseline by PsARC response assessed in each of the OPAL trials. [redacted]

5.2.8 Resources and costs

The CS provided a detailed description of resource use and costs incurred in PsA patients. These included: drug acquisition costs (Section B.3.5.2.2 in CS); drug administration costs (Section B.3.5.2.3 in CS) and drug monitoring costs (Section B.3.5.2.4). AEs costs were not considered in the model. A systematic review was conducted to identify alternative evidence regarding resource use and the costs associated with the management of PsA in the UK. The company reports that they did find one publication, Poole et al [35], that specifically reported estimates of costs according to HAQ-DI and/or PASI which was eligible for inclusion [35], however, it was not used to inform the model. The CS does not justify why this was not included. In TA445, HAQ-DI and PASI costs were based on the

same function as used in the York model (TA199) rather than the costs reported by Poole et al [35] TA445 concluded that this was due to limitations in the Poole et al study and to ensure consistency across NICE TAs.

Costs for acquisition, administration and monitoring differ between the first cycle (initiation phase) and subsequent cycles to reflect clinical management practices associated with switching a patient onto a new treatment. In addition, in the first cycle, monitoring is more intensive while the decision to continue with treatment is made. For comparators with a recommended initiation phase greater than 12 weeks (ustekinumab and sekukinumab), costs for the SPC recommended length of initiation phase were applied, for example up to 24 weeks. For other comparators the first cycle incorporates 12 weeks of drug treatment.

Table 46 in the CS (p138) provides a table detailing a summary of the treatment costs.

5.2.8.1 Drug acquisition costs

Costs for the bDMARDs and apremilast were sourced from the British National Formulary [19] and the cost of methotrexate was obtained from the electronic market information tool (eMIT) database [20]. PAS prices were used in the model where information is in the public domain. A list price analysis for tofacitinib was not provided. Instead the PAS price which employs a simple discount was used. Since the submission of the manuscript, the company have provided an updated PAS price for tofacitinib (See confidential PAS appendix). List prices were used for secukinumab and apremilast but these are subject a confidential PAS. The ERG conducted additional analysis using PAS prices for secukinumab and apremilast and these are presented in a confidential appendix. Biosimilar prices were used when available (etanercept and infliximab). No drug costs were assumed for BSC as it was assumed that these drug costs are captured in the estimates of resource use associated with HAQ-DI.

Following the update on the PAS price for tofacitinib, the company submitted a PAS submission template including tables detailing the new ICER using the confidential PAS price. They also provided an updated version of the model for each sub-population including the new PAS price. In sub-population 3, the incremental ICERs reported in the submission do not correspond to the incremental ICERs in the base case results in the economic model.

5.2.8.2 Drug administration costs

Administration costs were taken from the NHS reference costs [21, 22].

An intravenous infusion cost of £241 [21] is applied in each cycle for infliximab. This value is a weighted average cost for simple parenteral chemotherapy at first attendance, taking into account day case, outpatient and other costs, taken from NHS reference costs [21] as per TA445 [15]. For treatments

that require administration by subcutaneous injection, the cost of one hour of hospital-based nurse specialist time is applied (£45) to reflect clinical practice for bDMARDs prescribed by rheumatologists [22]. This cost is implemented in the first cycle only as it is assumed that the patient will self-administer subsequent treatment following training by the nurse.

The company did not assign a resource use associated with the administration of tofacitinib, apremilast or csDMARDs as these are taken orally.

5.2.8.3 Drug monitoring costs

Monitoring activities included in the model and their frequency of use (Table 45) are based on the assumptions from TA199 [14] and TA445 [15].

In the first cycle, patients undergo tests – full blood count, erythrocyte sedimentation rate, liver function test and urea and electrolytes – at the start of treatment and at month 3. In subsequent cycles, these tests are conducted every 6 months. The chest x-ray, tuberculosis Heaf test, antinuclear antibody, double-stranded DNA test and specialist visit are assumed to occur in the first cycle only.

Costs were taken from NHS reference costs [21], except for the liver function test, chest x-ray and tuberculosis Heaf test costs, which were inflated from the costs presented by the AG in TA445 [15].

The company assumes that the monitoring of tofacitinib is not considered as additional to current practice, and is in line with NHS policy for bDMARDs. However, in Table 3 of the CS (page 18), the company reports that tofacitinib monitoring requirements and identifies lipid testing at 8 weeks after commencing treatment. This monitoring requirement for tofacitinib was also identified by the ERG's clinical advisor. This is not included as a monitoring cost in the economic model.

In addition, as an oral therapy taken twice daily, patient adherence to tofacitinib may be an issue that would also justify additional monitoring. As additional monitoring or testing is likely to be of minimal cost (based on blood test costs in Table 45, page 136 in the CS), the ERG do not deem it necessary to explore this further in Section 6.

5.2.8.4 Disease related costs

In addition to drug acquisition, administration and monitoring costs, disease-related costs were also incorporated into the economic model.

Arthritis-related costs were estimated as a function of HAQ-DI score (Equation 5.1). For the model presented in this submission the annual direct cost was calculated using the formula from Rodgers et al [36], with costs inflated to 2017 prices:

Equation 5.2 Arthritis annual direct cost

$$Annual\ direct\ cost = £466.47\ x\ HAQ + £1{,}547.04$$

With the exception of BSC, these costs incorporate a 15% reduction to account for drug costs, in accordance with the York PsA model [14]. This is not applied to BSC as drug costs are assumed to be captured within health state costs and are not applied separately. This accords with the approach used in TA445[15].

The psoriasis component of resource use has previously been estimated based on PASI scores. Costs associated with the psoriasis component based on PASI scores were taken from the AG report in TA445 [15] and inflated to 2017 prices.

This analysis follows the approach taken in the York PsA model in TA199 [14] and TA445 [15]. The AG estimated costs for patients receiving bDMARDs based on baseline severity of psoriasis and whether or not they had a PASI75 response. For patients with mild–moderate or moderate–severe psoriasis at baseline achieving a PASI 75 response, the monthly estimated cost of a patient in remission [37] was applied. The source of this cost is a study which considered the cost-effectiveness of an intervention for patients with moderate to severe psoriasis in a Dutch setting. Costs from this analysis were similar to NHS reference costs and the company argues that the Dutch costs were generalizable to the UK after currency conversion.

Patients with moderate to severe psoriasis not achieving a PASI75 response were assumed to undergo one course of ultraviolet B treatment (UVB) per year. This incorporated the cost of the initial course of treatment and the cost of follow-up for the year. Patients were put into three categories for response – no response, response maintained for 12 months, and response maintained for 6 months followed by relapse. The total cost for the year was weighted by the frequency of these outcomes in the Hartman analysis (2002) [37].

Patients with mild to moderate psoriasis and no PASI 75 response were also assumed to receive a course of UVB but with the cost taken from NHS reference costs. The proportion of responders was taken from an analysis by Poyner et al (1999) (197). Patients with no baseline psoriasis incurred no costs.

5.2.8.5 Adverse reaction unit costs and resource use

As discussed in Section 5.2.6.6, AEs were not included explicitly in the model, neither as a treatment-related utility decrement nor as additional cost for treatment of adverse events. The company stated in their submission that this is consistent with previous TAs [15].

Adverse event costs were not explicitly included in the cost-effectiveness analysis; however, they influence response probabilities and withdrawal rates. This is in line with the approach used in previous models [15].

5.2.9 Base case cost effectiveness results

The following base case cost effectiveness results are the updated results provided by the company following clarification (including the updated PAS price for tofacitinib and corrected NMA results).

The expected costs and QALYs of the alternative treatments are reported for each sub-population and the relative cost-effectiveness of each strategy is compared using standard decision rules, estimating ICERs as appropriate. The base case analysis considers PAS prices for tofacitinib (updated PAS price) and its comparators, where these PAS prices are publicly available; certolizumab, golimumab and ustekinumab. Biosimilar prices are used for etanercept and infliximab. List prices were used for two products for which PAS schemes are approved but not publicly available; secukinumab and apremilast. The ERG conducted further analysis using the confidential PAS schemes for apremilast and secukinumab and these are presented in a separate confidential appendix.

5.2.9.1 People whose disease has not responded adequately to at least 2 non-biological DMARDs

The ICER for tofacitinib vs BSC (Table 32) is £13,419 per QALY. This result indicates that the inclusion of tofacitinib as an additional line of treatment for this sub-population falls within acceptable WTP thresholds as defined by NICE (between £20,000 and £30,000 per QALY gained). Based on the full incremental analysis, a strategy commencing with etanercept offers higher QALYs and falls within acceptable thresholds.

Table 32 Base case analysis (sub-population 2) (Table 8, p16 of PAS Template)

Strategy	Total discounted costs	Total discounted QALYs	Incremental cost vs. cheapest strategy	Incremental QALYs vs. cheapest strategy	ICER vs. cheapest strategy	Incremental ICER
BSC	[redacted]	[redacted]	-	-	-	-
TOF→UST→BSC	[redacted]	[redacted]	£32,881	2.45	£13,419	£13,419
APR→UST→BSC	[redacted]	[redacted]	£40,499	2.07	£19,569	Dominated
ADA→UST→BSC	[redacted]	[redacted]	£47,901	2.71	£17,687	Extendedly dominated
CTZ→UST→BSC	[redacted]	[redacted]	£48,839	2.85	£17,126	Extendedly dominated
ETN→UST→BSC	[redacted]	[redacted]	£51,700	3.27	£15,798	£22,886
SEK→UST→BSC	[redacted]	[redacted]	£52,978	2.86	£18,543	Dominated
GOL→UST→BSC	[redacted]	[redacted]	£53,557	2.99	£17,904	Dominated
INF→UST→BSC	[redacted]	[redacted]	£71,190	3.35	£21,225	£239,101

Abbreviations: ICER, incremental cost-effectiveness ratio; PAS, patient access scheme; QALY, quality-adjusted life-year.

5.2.9.2 People whose disease has not responded adequately to non-biological DMARDs and one or more TNFis

The ICER for tofacitinib 5 mg BD vs BSC (Table 33) is £9,001 per QALY. This result indicates that the inclusion of tofacitinib as an additional line of treatment for this sub-population falls within acceptable WTP thresholds as defined by NICE (between £20,000 and £30,000 per QALY gained). Based on the incremental analysis, a strategy commencing with tofacitinib is the only strategy that falls within an acceptable threshold.

Table 33 Base case analysis (sub-population 3) (Table 10, p17 of PAS Template)

Strategy	Total discounted costs	Total discounted QALYs	Incremental cost vs. cheapest strategy	Incremental QALYs vs. cheapest strategy	ICER vs. cheapest strategy	Incremental ICER
BSC	[redacted]	[redacted]	-	-	-	-
TOF→BSC	[redacted]	[redacted]	£11,732	1.30	£9,001	£9,001
UST→BSC	[redacted]	[redacted]	£26,709	1.42	£18,761	£124,510
SEC→BSC	[redacted]	[redacted]	£54,206	1.60	£33,914	£157,429

Abbreviations: ICER, incremental cost-effectiveness ratio; PAS, patient access scheme; QALY, quality-adjusted life-year.

5.2.9.3 People in whom TNFis are contraindicated or not tolerated

The ICER for tofacitinib 5 mg BD vs BSC (Table 34) is £7,825 per QALY. Similar to the previous results, this indicates that the inclusion of tofacitinib as an additional line of treatment for this sub-population falls within acceptable WTP thresholds as defined by NICE (between £20,000 and £30,000 per QALY gained). Based on the incremental analysis and similar to sub-population 3, a strategy commencing with tofacitinib is the only strategy that falls within an acceptable threshold.

Table 34 Base case analysis (sub-population 4) (Table 12, p 18 of PAS Template)

Strategy	Total discounted costs	Total discounted QALYs	Incremental cost vs. cheapest strategy	Incremental QALYs vs. cheapest strategy	ICER vs. cheapest strategy	Incremental ICER
BSC	[redacted]	[redacted]	-	-	-	-
TOF→BSC	[redacted]	[redacted]	£8,930	1.14	£7,825	£7,825
UST→BSC	[redacted]	[redacted]	£24,979	1.33	£18,837	Extendedly dominated
SEK→BSC	[redacted]	[redacted]	£30,153	1.62	£18,557	£43,872

Abbreviations: ICER, incremental cost-effectiveness ratio; PAS, patient access scheme; QALY, quality-adjusted life-year.

5.2.9.4 Conclusion on cost-effectiveness results

In each of the three sub-populations assessed, the deterministic ICER for tofacitinib 5 mg BD vs BSC was below £20,000 per QALY. In terms of the incremental analysis, a strategy commencing with tofacitinib is the only strategy that falls within an acceptable threshold in sub-populations 3 and 4,

whilst in sub-population 2, in addition to tofacitinib, etanercept provides higher QALYs whilst also falling within the acceptable threshold.

The cost-effectiveness results may, however, be sensitive to a number of assumptions made in the model, namely the choice of the NMA model and the PAS drug cost included for tofacitinib compared to the list prices incorporated for the other comparators for which PAS schemes are available but CiC, e.g. apremilast and secukinumab. The impact of these assumptions on cost-effectiveness is addressed by the ERG in Section 6.3 and in a separate confidential appendix.

As the economic model is similar to TA445 and included similar treatment comparators, the ERG have compared the costs and QALYs of the cost-effectiveness results with those in TA445. Given the difference in the psoriasis groups, this comparison is problematic so the ERG compared the average costs and QALYs across the psoriasis sub-groups in TA445 to compare with the current TA. The ERG conclude that the costs and QALYS between both TAs are relatively similar for each treatment. (See Appendix B in section 10)

5.2.10 Sensitivity analysis

The company presented a series of probabilistic sensitivity analysis (PSA) to assess the implications of parameter uncertainty, in terms of the estimates of cost-effectiveness. All parameters were assigned distributions and varied jointly. Ten thousand Monte Carlo simulations were recorded. Scatter plots and cost-effectiveness acceptability curves for the three sub-populations were presented in the CS (Figures 23-28 in CS).

The average results of PSA in all three sub-populations were consistent with the deterministic analyses and demonstrate that the ICER for tofacitinib 5 mg BD sequence remains below a threshold of £20,000 per QALY versus BSC in all sub-populations, where parameter uncertainty is explored.

In sub-population 2, the ICER versus BSC for the tofacitinib 5 mg BD sequence was only second to etanercept biosimilar, and in sub-populations 3 and 4, the tofacitinib 5 mg BD sequence was associated with the highest probability of being cost-effective at conventional willingness to pay thresholds of £20,000 and £30,000 per QALY.

Given that the probabilistic results in each of the sub-populations are similar to the results described in the deterministic analysis, Section 5.2.9, the ERG concludes that there are no particular concerns regarding non-linearity in the model.

5.2.12 Scenario Analysis

The CS included a series of scenario analysis that were performed to check the robustness of the model to structural assumptions made in the model The scenarios that were investigated along with a brief description of the assumptions for each are provided below.

List price analysis

- An alternative scenario using the list price of tofacitinib was considered.

Pessimistic NMA

- Alternate NMAs with worst outcomes for tofacitinib only were implemented to present a lower bound on the NMA analysis

Optimistic NMA

- Alternate NMAs with best outcome for tofacitinib only were used to present an upper bound on the NMA analysis.

ACR20 stopping rules

- To test the assumption of the PsARC stopping rules, response was defined by ACR20 response.

Pfizer mapping algorithm for all treatments

- To allow population-specific prediction of utility, the Pfizer mapping algorithm was applied instead of the algorithm from TA199 [14].

Pfizer mapping algorithm for tofacitinib only

- To allow population-specific prediction of utility, the Pfizer mapping algorithm is applied to the tofacitinib arm only.

The scenario analysis showed that tofacitinib 5 mg BD falls below (or between) the conventional NICE threshold of £20,000 to £30,000 per QALY across a range of plausible settings for all sub-populations (Tables 57 to 61 in CS). The results of the scenario analysis are consistent with the results presented in the base case analysis.

5.2.11 Model validation and face validity check

The CS reports that the cost-effectiveness model was validated by the model developers and by health economists not involved in the construction of the model. Validation was completed using standard procedures such as; cell-by-cell checks of logic and consistency, logical check of model outputs, and comparing outputs to those from previous economic analyses. The company did not provide specific details of the validation conducted and if the model failed on any aspects of the validation.

The ERG identified discrepancies in several of the efficacy results between those reported in the clinical section of the submission and the values that were subsequently used in the economic model.

The code for the PSA is complex and difficult to validate. In order to validate the model, the ERG requested the following;

- A step-by-step description of how the VBA code implements the PSA, including how the Monte Carlo is implemented.
- Confirmation of whether the simulations are done simultaneously for all comparators or separately for each individual comparator.
- Detailed annotations within the VBA code for each step.

In their response to clarification, the company confirmed that the simulations are performed simultaneously for all comparators and the company provided a detailed response on how the VBA code implements the PSA. They also provided additional annotation of the VBA code in the updated versions of the electronic models. Following the company response the ERG were able to validate the PSA and confirm that the PSA was conducted appropriately.

5.3 Exploratory and sensitivity analyses undertaken by the ERG

The ERG checked the model for consistency with the CS. In consistencies are detailed in the sections above. The ERG also checked the company model for any errors and validated assumptions noted in the CS.

Sensitivity analyses conducted by the ERG are detailed in Section 6.

5.4 Conclusions of the cost effectiveness section

The ERG has a number of concerns regarding some of the approaches, assumptions and data used in the CS and economic model. The main concerns expressed by the ERG are:

The PsARC response data used by the company in the optimistic and pessimistic scenarios may not reflect the best fitting NMA model. There is also no clear rationale for the placebo effect, and hence that the results of the placebo-response adjusted model should be interpreted with caution. We will therefore explore the use of use the independent treatment effects model and the class effect model proposed by the in sensitivity analysis in Section 6. The corrected errors in the company preferred model (B2) are also propagated through the company model in Section 6.

Disease progression

As stated in Section 5.2.1, the company assumes that HAQ-DI progression stops when a patient responds to tofacitinib. The ERG is concerned about this assumption given that there is no long-term evidence on radiographic progression on tofacitinib to support this assumption. In section 6, the ERG addresses this assumption by conducting scenario analyses using different rates of HAQ-DI progression. The first scenario considers the impact of tofacitinib progression equal to that of apremilast. The second scenario considers radiographic progression reported in the adalimumab study [33] where 11% of patients progressed on treatment. Finally, the ERG considers the impact on cost-effectiveness if 11% of the population progress at the same rate as assumed for apremilast (0.010). The rates used in the scenario analysis are reflective of the radiographic progression study referred to in the CS [33].

Psoriasis sub-groups

As stated in Section 5.2.3, the sub-populations in the model are not defined according to psoriasis level and the ERG have concerns about this assumption given the impact that differences in baseline characteristics such as HAQ-DI, and particularly PASI scores can have on cost-effectiveness results. This is an issue in terms of the severity of psoriasis and the consequent dosing of comparators such as secukinumab; where secukinumab 300mg is approved for patients with severe psoriasis as opposed to the standard does of secukinumab 150mg. The ERG considers the impact of defining the sub-populations by psoriasis level to reflect the approach taken previously in TA445.

Drug acquisition costs

The company used PAS prices that were publicly available, namely for ustekinumab and certolizumab. Biosimilar costs were assumed for infliximab and etanercept. For other comparators the list price of the drug was implemented in the model. The ERG has concerns regarding the impact of other PAS schemes, apremilast and secukinumab and the impact that this may have on the cost-effectiveness of tofacitinib. The ERG considers the impact of including the PAS prices for APR and SEC in the confidential appendix.

Effect degradation for subsequent lines of therapy

The CS does not apply a reduction in effectiveness for subsequent lines of therapy. As discussed in Section 5.2.6.2, this may overestimate the cost-effectiveness of treatments with a lower PsARC response rate. In TA445 the effect degradation was estimated from observational data for RA patients from the BSR register. For a patient that failed first line therapy due to lack of efficacy, the risk of failing the second-line therapy due to lack of efficacy increases by 2.7 (95% CI 2.1-3.4). The ERG is

unable to explore the sensitivity of the cost-effectiveness results to this assumption, due to the inflexibility of the company model provided.

Given the importance of the issues discussed, additional analyses undertaken by the ERG are presented in Section 6, which consider the potential impact of these uncertainties on the cost-effectiveness results.

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

6.1 Overview

This section details the ERG's further exploration of the assumptions and uncertainties raised in the review and critique of the manufacturer's cost effectiveness analysis, presented in Section 5. The ERG present alternative estimates of cost effectiveness, correcting the company model and also explore assumptions and data incorporated in the manufacturer's analysis.

The ERG's exploratory analyses focused on the following key issue and uncertainty:

- NMA sensitivity analyses with PsARC corrected base case model in sub-population 2. (Section 4.7.1).

Additional scenarios around specific model parameters were:

- Severity of psoriasis by subgroup.
- Tofacitinib progression rates for PsARC responders.
- Drug costs for comparator drugs. The drug cost analyses are based on PAS schemes that are approved for Secukinumab and Apremilast but are not in the public domain. The results of the additional analyses by each sub-population are included in a separate confidential appendix.

These scenarios are meant to be exploratory in nature and are intended to show the impact of different parameter assumptions on the cost-effectiveness results. The ICERs for all scenarios are presented both compared to the cheapest strategy (BSC) and as a fully incremental analysis. For the ERG conducted sensitivity analyses, both deterministic and probabilistic results are presented. For the additional scenarios, deterministic results are presented using the most valid NMA model concluded in Section 6.2.1.

6.2 ERG corrections and adjustments to the company's base case model

6.2.1 NMA Sensitivity analysis

In Section 4.7.1 the ERG present a corrected NMA for B2, the placebo adjusted random effects model, generating alternative estimates of PsARC response for subpopulation 2 (see Table 23). In addition the ERG explored the use of a class effects model (D2) and an independent treatment effects model (A2), concluding that D2 represents the model with the best fit (lowest DIC). The ERG also conclude that the placebo adjusted model should be interpreted with caution, due to a lack of rationale for the placebo effect.

In this section the ERG explore the sensitivity of the company cost-effectiveness results to alternative NMA models to estimate PsARC response rates, specifically the corrected B2, D2 and A2 for subpopulation 2. The results for these sensitivity analyses are presented below in Table 35 to Table 42, for both the deterministic and the probabilistic analysis. The equivalent confidential PAS results are presented in a separate confidential appendix.

Table 35 Company base case results B2 (deterministic)

Strategy	Total discounted costs	Total discounted QALYs	Incremental cost vs. cheapest strategy	Incremental QALYs vs. cheapest strategy	ICER vs. cheapest strategy	Incremental ICER
BSC	[redacted]	[redacted]	-	-	-	-
TOF	[redacted]	[redacted]	£32,881	2.45	£13,419	£13,419
APR	[redacted]	[redacted]	£40,499	2.07	£19,569	Dominated
ADA	[redacted]	[redacted]	£47,901	2.71	£17,687	Extendedly dominated
CTZ	[redacted]	[redacted]	£48,839	2.85	£17,126	Extendedly dominated
ETN	[redacted]	[redacted]	£51,700	3.27	£15,798	£22,886
SEK	[redacted]	[redacted]	£52,978	2.86	£18,543	Dominated
GOL	[redacted]	[redacted]	£53,557	2.99	£17,904	Dominated
INF	[redacted]	[redacted]	£71,190	3.35	£21,225	£239,101

Table 36 Company base case results B2 (probabilistic)

Strategy	Total discounted costs	Total discounted QALYs	Incremental cost vs. cheapest strategy	Incremental QALYs vs. cheapest strategy	ICER vs. cheapest strategy	Incremental ICER
BSC	[redacted]	[redacted]	-	-	-	-
TOF	[redacted]	[redacted]	£33,231	2.39	£13,918	£13,918
APR	[redacted]	[redacted]	£40,841	2.00	£20,422	Dominated
ADA	[redacted]	[redacted]	£48,350	2.64	£18,318	Extendedly dominated
CTZ	[redacted]	[redacted]	£49,313	2.77	£17,815	Extendedly dominated
ETN	[redacted]	[redacted]	£52,182	3.19	£16,371	£23,696
SEK	[redacted]	[redacted]	£53,510	2.78	£19,253	Dominated
GOL	[redacted]	[redacted]	£54,009	2.90	£18,641	Dominated
INF	[redacted]	[redacted]	£71,630	3.27	£21,900	£233,602

Table 37 ERG B2 – base case results (deterministic)

Strategy	Total discounted costs	Total discounted QALYs	Incremental cost vs. cheapest strategy	Incremental QALYs vs. cheapest strategy	ICER vs. cheapest strategy	Incremental ICER
BSC	[redacted]	[redacted]	-	-	-	-
TOF	[redacted]	[redacted]	£32,822	2.52	£13,029	£13,029
APR	[redacted]	[redacted]	£39,434	2.02	£19,555	Dominated
ADA	[redacted]	[redacted]	£47,275	2.67	£17,701	Extendedly dominated
CTZ	[redacted]	[redacted]	£49,490	2.89	£17,145	Extendedly dominated
ETN	[redacted]	[redacted]	£50,598	3.20	£15,799	£26,006
GOL	[redacted]	[redacted]	£51,143	2.85	£17,931	Dominated
SEK	[redacted]	[redacted]	£53,774	2.91	£18,507	Dominated
INF	[redacted]	[redacted]	£69,389	3.26	£21,270	£315,590

Table 38 ERG B2 – base case results (probabilistic)

Strategy	Total discounted costs	Total discounted QALYs	Incremental cost vs. cheapest strategy	Incremental QALYs vs. cheapest strategy	ICER vs. cheapest strategy	Incremental ICER
BSC	[redacted]	[redacted]	-	-	-	-
TOF	[redacted]	[redacted]	£33,231	2.39	£13,918	£13,0244
APR	[redacted]	[redacted]	£40,841	2.00	£20,422	Dominated
ADA	[redacted]	[redacted]	£48,350	2.64	£18,318	Extendedly dominated
CTZ	[redacted]	[redacted]	£49,313	2.77	£17,815	Extendedly dominated
ETN	[redacted]	[redacted]	£52,182	3.19	£16,371	£25,762
GOL	[redacted]	[redacted]	£54,009	2.90	£18,641	Dominated
SEK	[redacted]	[redacted]	£53,510	2.78	£19,253	Dominated
INF	[redacted]	[redacted]	£71,630	3.27	£21,900	£216,088

The corrected B2 NMA produces very similar results to the company base case results, with only small differences in costs and QALYs and ICERs compared to BSC and the full incremental. The deterministic and probabilistic versions also provide similar results in terms of ordering, although there are some discrepancies in terms of absolute costs and QALYs. For all comparators the ICERs versus BSC fall within acceptable thresholds for cost-effectiveness. For the company B2 model and the corrected B2 model, both tofacitinib and etanercept fall within acceptable thresholds for the full incremental analysis.

Table 39 ERG D –base case results (deterministic)

Strategy	Total discounted costs	Total discounted QALYs	Incremental cost vs. cheapest strategy	Incremental QALYs vs. cheapest strategy	ICER vs. cheapest strategy	Incremental ICER

BSC	[redacted]	[redacted]	-	-	-	-
TOF	[redacted]	[redacted]	£34,099	2.62	£13,011	£13,011
APR	[redacted]	[redacted]	£40,487	2.07	£19,533	Dominated
ADA	[redacted]	[redacted]	£48,963	2.77	£17,665	Extendedly dominated
CTZ	[redacted]	[redacted]	£50,481	2.95	£17,138	Extendedly dominated
ETN	[redacted]	[redacted]	£50,635	3.19	£15,855	£28,866
GOL	[redacted]	[redacted]	£51,798	2.89	£17,911	Dominated
SEK	[redacted]	[redacted]	£54,680	2.96	£18,476	Dominated
INF	[redacted]	[redacted]	£68,835	3.25	£21,176	£320,148

Table 40 ERG D –base case results (probabilistic)

Strategy	Total discounted costs	Total discounted QALYs	Incremental cost vs. cheapest strategy	Incremental QALYs vs. cheapest strategy	ICER vs. cheapest strategy	Incremental ICER
BSC	[redacted]	[redacted]	-	-	-	-
TOF	[redacted]	[redacted]	£34,514	2.55	£13,530	£13,529
APR	[redacted]	[redacted]	£40,870	2.01	£20,310	Dominated
ADA	[redacted]	[redacted]	£49,520	2.71	£18,276	Extendedly dominated
CTZ	[redacted]	[redacted]	£51,200	2.88	£17,789	Extendedly dominated
ETN	[redacted]	[redacted]	£51,317	3.13	£16,414	£29,199
GOL	[redacted]	[redacted]	£52,258	2.81	£18,601	Dominated
SEK	[redacted]	[redacted]	£55,277	2.89	£19,156	Dominated
INF	[redacted]	[redacted]	£69,735	3.20	£21,801	£255,288

Again the D2 NMA produces very similar results to the company base case results and the correct B2 results. There are only small differences in costs and QALYs and ICERs compared to BSC and the full incremental. The deterministic and probabilistic versions also provide similar results, suggesting that the ICERs for all comparators versus BSC fall within acceptable thresholds for cost-effectiveness. In the full incremental the ICERs for both tofacitinib and etanercept both fall within acceptable thresholds.

Table 41 ERG A2 - base case results (deterministic)

Strategy	Total discounted costs	Total discounted QALYs	Incremental cost vs. cheapest strategy	Incremental QALYs vs. cheapest strategy	ICER vs. cheapest strategy	Incremental ICER

BSC	[redacted]	[redacted]	-	-	-	-
TOF	[redacted]	[redacted]	£29,255	2.19	£13,355	£13,355
APR	[redacted]	[redacted]	£37,505	1.91	£19,664	Dominated
ADA	[redacted]	[redacted]	£44,565	2.51	£17,771	Extendedly dominated
CTZ	[redacted]	[redacted]	£44,690	2.61	£17,151	Extendedly dominated
SEK	[redacted]	[redacted]	£48,122	2.56	£18,765	Dominated
ETN	[redacted]	[redacted]	£49,290	3.14	£15,716	£21,186
GOL	[redacted]	[redacted]	£52,253	2.91	£17,959	Dominated
INF	[redacted]	[redacted]	£70,233	3.27	£21,480	£156,878

Table 42 ERG A2 base case results (probabilistic)

Strategy	Total discounted costs	Total discounted QALYs	Incremental cost vs. cheapest strategy	Incremental QALYs vs. cheapest strategy	ICER vs. cheapest strategy	Incremental ICER
BSC	[redacted]	[redacted]				
TOF	[redacted]	[redacted]	£29,780	2.11	£14,109	£14,109
APR	[redacted]	[redacted]	£38,027	1.82	£20,871	Dominated
CTZ	[redacted]	[redacted]	£45,149	2.50	£18,031	Extendedly dominated
ADA	[redacted]	[redacted]	£45,177	2.42	£18,654	Dominated
SEK	[redacted]	[redacted]	£48,633	2.47	£19,713	Dominated
ETN	[redacted]	[redacted]	£49,936	3.04	£16,448	£21,782
GOL	[redacted]	[redacted]	£52,840	2.80	£18,874	Dominated
INF	[redacted]	[redacted]	£70,781	3.17	£22,336	£156,769

The A2 NMA produces very similar results to the company base case results and the correct B2 results in terms of costs and QALYs. The QALYs are however consistently lower for all comparators compared to the B2 and D2 models. This is expected due to the lower PsARC response rates predicted using the independent treatment effects model (A2) compared with the placebo adjusted models (B2 and D2) (see Table 23).

There are only small differences in the ICERs compared to BSC, suggesting that the ICERs for all comparators versus BSC fall within acceptable thresholds for cost-effectiveness. The full incremental ICERs also show that etanercept and tofacitinib are likely to fall within acceptable ranges for the threshold and these are lower than the ICERs for the B2 and D2 NMA models. The deterministic and probabilistic versions provide similar results.

6.3 Additional ERG analyses

6.3.1 Severity of psoriasis

As discussed in Section 5.2.3, the sub-populations were not defined according to psoriasis level as specified in TA445. Instead, a weighted average PASI score of the psoriasis subgroups was calculated

for the entire population. The ERG had concerns about this assumption given the impact that differences in baseline characteristics such as HAQ-DI, and particularly PASI scores can have on cost-effectiveness results and the appropriateness of some comparators for particular levels of psoriasis. In particular, different secukinumab dosages are appropriate for the separate sub-populations: 150mg of SEC for naïve patients without psoriasis or with mild to moderate psoriasis and 300mg of SEC for experienced patients and for naïve patients with moderate to severe psoriasis[15]. The company model assumes a SEC weighted dose for sub-populations 2 and 4. For sub-population 3 the appropriate 300mg dose for secukinumab was applied in the company model. The ERG considered the impact of defining sub-populations 2 and 4 by psoriasis level and applying the appropriate dosage of secukinumab as described previously. This sensitivity analysis uses the ERG preferred NMA model (D2).

Table 43 Sub-population 2 defined by psoriasis level

NO PSORIASIS – SEK 150MG						
Strategy	**Total discounted costs**	**Total discounted QALYs**	**Incremental cost vs. cheapest strategy**	**Incremental QALYs vs. cheapest strategy**	**ICER vs. cheapest strategy**	**Incremental ICER**
BSC	[redacted]	[redacted]	-	-	-	-
TOF	[redacted]	[redacted]	£37,000	2.57	£14,400	£14,396
APR	[redacted]	[redacted]	£43,110	2.03	£21,272	Dominated
SEK	[redacted]	[redacted]	£51,072	2.89	£17,675	Extendedly dominated
ADA	[redacted]	[redacted]	£52,057	2.72	£19,165	Dominated
CTZ	[redacted]	[redacted]	£53,358	2.89	£18,433	Extendedly dominated
ETN	[redacted]	[redacted]	£53,417	3.14	£16,986	£28,530
GOL	[redacted]	[redacted]	£55,525	2.82	£19,658	Dominated
INF	[redacted]	[redacted]	£73,195	3.17	£23,076	£732,175
MILD TO MODERATE PSORIASIS – SEK150MG						
BSC	[redacted]	[redacted]	-	-	-	-
TOF	[redacted]	[redacted]	£34,115	2.65	£12,897	£12,896
APR	[redacted]	[redacted]	£40,501	2.09	£19,336	Dominated
SEK	[redacted]	[redacted]	£47,388	2.99	£15,859	Extendedly dominated
ADA	[redacted]	[redacted]	£48,979	2.80	£17,505	Dominated
CTZ	[redacted]	[redacted]	£50,497	2.97	£17,003	Dominated
ETN	[redacted]	[redacted]	£50,650	3.22	£15,745	£28,925
GOL	[redacted]	[redacted]	£51,818	2.92	£17,722	Dominated
INF	[redacted]	[redacted]	£68,859	3.29	£20,943	£256,411
MODERATE TO SEVERE – SEK300MG DOSE						
BSC	[redacted]	[redacted]	-	-	-	-
TOF	[redacted]	[redacted]	£28,282	2.70	£10,477	£10,477

APR	[redacted]	[redacted]	£35,227	2.14	£16,438	Dominated
SEK	[redacted]	[redacted]	£69,046	3.11	£22,187	Dominated
ADA	[redacted]	[redacted]	£42,757	2.86	£14,970	Extendedly dominated
CTZ	[redacted]	[redacted]	£44,711	3.02	£14,789	Extendedly dominated
ETN	[redacted]	[redacted]	£45,056	3.27	£13,786	£29,483
GOL	[redacted]	[redacted]	£44,323	2.99	£14,801	Extendedly dominated
INF	[redacted]	[redacted]	£60,091	3.37	£17,828	£146,891

In sub-population 2, the ICERs for tofacitinib in each psoriasis sub-group fall below the conventional NICE threshold range of £20,000 to £30,000 per QALY (Table 43). Based on the fully incremental analysis (also Table 43), a strategy commencing with etanercept is more effective (i.e. offers higher QALYs) than tofacitinib and falls within acceptable NICE thresholds.

Similarly, in sub-population 4, the tofacitinib ICERs in each psoriasis sub-group fall below the acceptable NICE thresholds (Table 44). In the *no psoriasis* and *moderate to severe* sub-group, tofacitinib is the only treatment with an ICER that does not exceed that of the NICE threshold. However, in the *mild to moderate* psoriasis sub-group, secukinumab offers higher QALYs than tofacitinib and lies just below the NICE acceptable threshold of £30,000.

Table 44 Sub-population 4 defined by psoriasis level

NO PSORIASIS – SEK 150MG						
Strategy	**Total discounted costs**	**Total discounted QALYs**	**Incremental cost vs. cheapest strategy**	**Incremental QALYs vs. cheapest strategy**	**ICER vs. cheapest strategy**	**Incremental ICER**
BSC	[redacted]	[redacted]	-	-	-	-
TOF	[redacted]	[redacted]	£10,068	1.12	£8,972	£8,972
SEK	[redacted]	[redacted]	£25,274	1.59	£15,936	£32,789
UST	[redacted]	[redacted]	£26,467	1.30	£20,353	Dominated
MILD TO MODERATE PSORIASIS – SEK150MG						
BSC	[redacted]	[redacted]	-	-	-	-
TOF	[redacted]	[redacted]	£8,936	1.15	£7,769	£7,769
SEK	[redacted]	[redacted]	£23,246	1.64	£14,181	£29,262
UST	[redacted]	[redacted]	£24,987	1.34	£18,671	Dominated
MODERATE TO SEVERE PSORIASIS – SEK300MG						
BSC	[redacted]	[redacted]	-	-	-	-
TOF	[redacted]	[redacted]	£6,647	1.17	£5,680	£5,680
UST	[redacted]	[redacted]	£21,997	1.37	£16,112	Extendedly dominated
SEK	[redacted]	[redacted]	£45,795	1.69	£27,137	£75,660

6.3.2 Tofacitinib progression rates

As described in Section 5.2.6.3, the ERG had concerns regarding the rate of tofacitinib progression given the lack of long-term evidence on radiographic progression. To assess this assumption, the ERG conducted scenario analyses using different HAQ-DI progression rates for tofacitinib. The first scenario assesses the impact on cost-effectiveness when tofacitinib progression is equal to that of Apremilast. In addition, based on the progression rates reported for Adalimumab [33], the ERG also considers a scenario where 11% of the population progress at the BSC rate and another scenario where 11% of the population progress at the apremilast progression rate. This sensitivity analysis uses the ERG preferred NMA model (D2) and the weighted level of psoriasis as in the company base-case.

Table 45 Sub-population 2: Tofacitinib progression rate scenarios

TOFACITINIB PROGRESSION = APREMILAST PROGRESSION						
Strategy	**Total discounted costs**	**Total discounted QALYs**	**Incremental cost vs. cheapest strategy**	**Incremental QALYs vs. cheapest strategy**	**ICER vs. cheapest strategy**	**Incremental ICER**
BSC	[redacted]	[redacted]	-	-	-	-
TOF	[redacted]	[redacted]	£34,785	2.21	£15,706	£15,706
APR	[redacted]	[redacted]	£40,487	2.07	£19,533	Dominated
ADA	[redacted]	[redacted]	£48,963	2.77	£17,665	Extendedly dominated
CTZ	[redacted]	[redacted]	£50,481	2.95	£17,138	Extendedly dominated
ETN	[redacted]	[redacted]	£50,635	3.19	£15,855	£16,191
GOL	[redacted]	[redacted]	£51,798	2.89	£17,911	Dominated
SEK	[redacted]	[redacted]	£54,680	2.96	£18,476	Dominated
INF	[redacted]	[redacted]	£68,835	3.25	£21,176	£320,148
TOFACITINIB PROGRESSION: 11% PROGRESS AT BSC (TOF PROGRESSION UPDATED TO 0.002)						
BSC	[redacted]	[redacted]	-	-	-	-
TOF	[redacted]	[redacted]	£34,251	2.53	£13,531	£13,531
APR	[redacted]	[redacted]	£40,487	2.07	£19,533	Dominated
ADA	[redacted]	[redacted]	£48,963	2.77	£17,665	Extendedly dominated
CTZ	[redacted]	[redacted]	£50,481	2.95	£17,138	Extendedly dominated
ETN	[redacted]	[redacted]	£50,635	3.19	£15,855	£24,735
GOL	[redacted]	[redacted]	£51,798	2.89	£17,911	Dominated
SEK	[redacted]	[redacted]	£54,680	2.96	£18,476	Dominated
INF	[redacted]	[redacted]	£68,835	3.25	£21,176	£320,148
TOFACITINIB: 11% PROGRESS AT SAME RATE AS APREMILAST (TOF PROGRESSION UPDATED TO 0.001)						
BSC	[redacted]	[redacted]	-	-	-	-
TOF	[redacted]	[redacted]	£34,175	2.58	£13,266	£13,266
APR	[redacted]	[redacted]	£40,487	2.07	£19,533	Dominated
ADA	[redacted]	[redacted]	£48,963	2.77	£17,665	Extendedly dominated

CTZ	[redacted]	[redacted]	£50,481	2.95	£17,138	Extendedly dominated
ETN	[redacted]	[redacted]	£50,635	3.19	£15,855	£26,650
GOL	[redacted]	[redacted]	£51,798	2.89	£17,911	Dominated
SEK	[redacted]	[redacted]	£54,680	2.96	£18,476	Dominated
INF	[redacted]	[redacted]	£68,835	3.25	£21,176	£320,148

Table 45 shows that for all progression scenarios, all comparators fall within the acceptable thresholds for cost-effectiveness, compared to BSC. For the fully incremental analysis for sub-population 2, a strategy commencing with etanercept offers higher QALYs in each scenario, however tofacitinib also has an ICER that falls below (or within) the acceptable NICE thresholds.

Table 46 Sub-population 3: Tofacitinib progression rate scenarios

TOFACITINIB PROGRESSION = APREMILAST PROGRESSION						
Strategy	**Total discounted costs**	**Total discounted QALYs**	**Incremental cost vs. cheapest strategy**	**Incremental QALYs vs. cheapest strategy**	**ICER vs. cheapest strategy**	**Incremental ICER**
BSC	[redacted]	[redacted]	-	-	-	-
TOF	[redacted]	[redacted]	£12,583	0.82	£15,400	£15,400
UST	[redacted]	[redacted]	£26,709	1.42	£18,761	£23,287
SEK	[redacted]	[redacted]	£54,206	1.60	£33,914	£157,429
TOFACITINIB: (11% PROGRESS AT BSC (TOF PROGRESSION UPDATED TO 0.002)						
BSC	[redacted]	[redacted]	-	-	-	-
TOF	[redacted]	[redacted]	£11,923	1.19	£9,984	£9,984
UST	[redacted]	[redacted]	£26,709	1.42	£18,761	£64,441
SEK	[redacted]	[redacted]	£54,206	1.60	£33,914	£157,429
TOFACITINIB: 11% PROGRESS AT SAME RATE AS APREMILAST (TOF PROGRESSION UPDATED TO 0.001)						
BSC	[redacted]	[redacted]	-	-	-	-
TOF	[redacted]	[redacted]	£11,828	1.25	£9,472	£9,472
UST	[redacted]	[redacted]	£26,709	1.42	£18,761	£85,041
SEK	[redacted]	[redacted]	£54,206	1.60	£33,914	£157,429

In sub-population 3 (Table 46), when the tofacitinib progression rate is equal to that of apremilast, ustekinumab offers higher QALYs and is associated with an ICER of £23,287. When 11% of patients progress at the same rate as BSC or apremilast, a strategy commencing with tofacitinib is the only strategy that falls within (below) the NICE acceptable threshold.

Table 47 Sub-population 4: tofacitinib progression rate scenarios

TOFACITINIB PROGRESSION = APREMILAST PROGRESSION						
Strategy	**Total discounted costs**	**Total discounted QALYs**	**Incremental cost vs. cheapest strategy**	**Incremental QALYs vs. cheapest strategy**	**ICER vs. cheapest strategy**	**Incremental ICER**
BSC	[redacted]	[redacted]	-	-	-	-

TOF	[redacted]	[redacted]	£9,655	0.73	£13,266	£13,266
UST	[redacted]	[redacted]	£24,979	1.33	£18,837	Extendedly dominated
SEK	[redacted]	[redacted]	£30,153	1.62	£18,557	£22,849
TOFACITINIB: (11% PROGRESS AT BSC (TOF PROGRESSION UPDATED TO 0.002)						
BSC	[redacted]	[redacted]	-	-	-	-
TOF	[redacted]	[redacted]	£9,092	1.05	£8,670	£8,670
UST	[redacted]	[redacted]	£24,979	1.33	£18,837	Extendedly dominated
SEK	[redacted]	[redacted]	£30,153	1.62	£18,557	£36,554
TOFACITINIB: 11% PROGRESS AT SAME RATE AS APREMILAST (TOF PROGRESSION UPDATED TO 0.001)						
BSC	[redacted]	[redacted]	-	-	-	-
TOF	[redacted]	[redacted]	£9,011	1.09	£8,230	£8,230
UST	[redacted]	[redacted]	£24,979	1.33	£18,837	Extendedly dominated
SEK	[redacted]	[redacted]	£30,153	1.62	£18,557	£39,888

Similar results for sub-population 4 are shown in Table 47 except for the first progression scenario (tofacitinib is equal to apremilast) where secukinumab offers higher QALYs and has an ICER within the NICE acceptable threshold.

6.4 Conclusions from ERG analyses

The ERG conducted a range of exploratory analyses to assess the uncertainties raised in the review and critique of the manufacturer's clinical and cost-effectiveness evidence. The ERG's exploratory analyses focussed on, severity of psoriasis, tofacitinib progression rates and drug costs for comparator drugs that are approved but not available publicly.

The additional analyses undertaken by the ERG suggested that whilst the ICERs for all subpopulations changed in each of the scenarios, they remained within the acceptable willingness to pay threshold, compared to BSC. In all scenarios, the fully incremental ICERs for tofacitinib are also within conventional willingness to pay thresholds, although etanercept may offer higher QALYs within an acceptable threshold. The confidential PAS appendix considers the impact of the PAS prices for apremilast and secukinumab on the cost-effectiveness results.

7 End of life

Not applicable.

8 Overall conclusions

The evidence for the clinical effectiveness of tofacitinib is based on good quality randomised trials and the results are likely to be reliable.

The ERG identified limitations in the generalisability of the RCT evidence to clinical practice:

- A significant proportion of patients in each RCT (18% and 24%) was treated in combination with sulfasalazine and leflunomide, when the marketing authorisation is for tofacitinib in combination with methotrexate (MTX) only.
- In both OPAL Broaden and OPAL Beyond the placebo-controlled phase was limited to 3 months: treatment with tofacitinib in clinical practice is long-term.
- The use of adalimumab in OPAL Broaden in combination with a csDMARD, is not reflective of adalimumab in clinical practice or in other trials;
- the number of previous TNFis (and the specific previous TNFis) in OPAL Beyond may not reflect the patient population in which tofacitnib will be used in current practice;
- and finally, in OPAL Balance (the long-term follow-up study) [redacted], whereas the licenced dose for tofacitinib in 5mg BD.

The ERG identified errors in the implementation of the company's placebo-adjusted NMAs. Models corrected by the ERG found a more meaningful interaction between baseline risk and treatment effect than the company analyses. The corrected ERG analyses also showed statistical support for models considering class-effects. Additionally, residual deviance for the placebo arm of OPAL BROADEN no longer indicated a poor fit. Therefore the ERG corrected models did not provide support for the company's 'optimistic analyses' (sensitivity analyses to improve goodness of fit) that excluded the placebo arm of OPAL Broaden.

The ERG had concerns regarding assumptions in the CS and economic model. In particular, the assumption that tofacitinib halts HAQ-DI progression while patients remain on treatment. The ERG is cautious of this assumption given that no long-term clinical evidence is available to support this, such as data assessing radiographic disease progression. The ERG assessed this assumption by conducting scenario analyses using different HAQ-DI progression rates. The ERG conclude that whilst the ICERs change for each sub-population, they remain within the acceptable willingness to pay threshold, compared to BSC. However, in the fully incremental analyses, tofacitinib is bettered by etanercept in sub-population 2 in each progression scenario. In sub-population 3, when tofacitinib progression is equal to apremilast progression, ustekinumab is more effective (i.e. offers more QALYs) than

tofacitinib. Similarly, in sub-population 4, when tofacitinib is equal to apremilast progression, secukinumab offers higher QALYs within an acceptable cost-effectiveness threshold.

The ERG also had concerns about assumptions made regarding effect degradation for subsequent lines of therapy. The CS does not apply a reduction in effectiveness for subsequent lines of therapy. This may over-estimate the cost-effectiveness of treatments with a lower PsARC response rate. Due to the lack of flexibility in the company model, the ERG is unable to explore the sensitivity of the cost-effectiveness results to this assumption.

8.1 Implications for research

Longer term data are required to confirm the efficacy of tofacitinib, particularly for the outcome of progression of joint disease and the implications this may have on cost-effectiveness.

9 References

1. Corbett M, Chehadah F, Biswas M, Moe-Byrne T, Palmer S, Soares M, et al. Certolizumab pegol and secukinumab for treating active psoriatic arthritis following inadequate response to disease-modifying antirheumatic drugs: a systematic review and economic evaluation. *Health Technol Assess* 2017;**21**:1-326.
2. Tillett W, Charlton R, Nightingale A, Snowball J, Green A, Smith C, et al. Interval between onset of psoriasis and psoriatic arthritis comparing the UK Clinical Practice Research Datalink with a hospital-based cohort. *Rheumatology (Oxford)* 2017;**56**:2109-13.
3. Husted JA, Thavaneswaran A, Chandran V, Gladman DD. Incremental effects of comorbidity on quality of life in patients with psoriatic arthritis. *J Rheumatol.* 2013;**40**:1349-56.
4. Gossec L, de Wit M, Kiltz U, Braun J, Kalyoncu U, Scrivo R, et al. A patient-derived and patient-reported outcome measure for assessing psoriatic arthritis: elaboration and preliminary validation of the Psoriatic Arthritis Impact of Disease (PsAID) questionnaire, a 13-country EULAR initiative. *Ann Rheum Dis* 2014;**73**:1012-9.
5. Lebwohl MG, Bachelez H, Barker J, Girolomoni G, Kavanaugh A, Langley RG, et al. Patient perspectives in the management of psoriasis: results from the population-based Multinational Assessment of Psoriasis and Psoriatic Arthritis Survey. *J Am Acad Dermatol* 2014;**70**:871.
6. Xu Y, Sudharshan L, Hsu MA, Koenig A, Cappelleri JC, Liu W, et al. *Patient preferences associated with the use of treatments for psoriatic arthritis: Results of a conjoint analysis [abstract]*. In: Arthritis Rheumatol.; 2017. Available from: http://acrabstracts.org/abstract/patient-preferences-associated-with-the-use-of-treatments-for-psoriatic-arthritis-results-of-a-conjoint-analysis/
7. Kimball AB, Okun M, Sundaram M, Mulani PM, Bai Y. Approved adalimumab dosing regimen associated with greater efficacy and lower cost per responder compared with 40 mg every other week dosing without initial 80 mg dose: analysis of outcomes from adalimumab psoriasis clinical trial database. *J Am Acad Dermatol* 2012;**66**:AB185.
8. Glintborg B, Ostergaard M, Dreyer L, Krogh NS, Tarp U, Hansen MS, et al. Treatment response, drug survival, and predictors thereof in 764 patients with psoriatic arthritis treated with anti-tumor necrosis factor alpha therapy: results from the nationwide Danish DANBIO registry. *Arthritis Rheum* 2011;**63**:382-90
9. Saad AA, Ashcroft DM, Watson KD, Hyrich KL, Noyce PR, Symmons DP. Persistence with anti-tumour necrosis factor therapies in patients with psoriatic arthritis: observational study from the British Society of Rheumatology Biologics Register. *Arthritis Res Ther* 2009;**11**:R52.
10. Fagerli KM, Kearsley-Fleet L, Watson KD, Packham J, BSRBR-RA Contributors Group, Symmons DP, et al. Long-term persistence of TNF-inhibitor treatment in patients with psoriatic arthritis. Data from the British Society for Rheumatology Biologics Register. *RMD Open* 2018;**4**:e000596.
11. Taylor W, Gladman D, Helliwell P, Marchesoni A, Mease P, Mielants H. Classification criteria for psoriatic arthritis: development of new criteria from a large international study. *Arthritis Rheum* 2006;**54**:2665-73.
12. Eder L, Thavaneswaran A, Chandran V, Gladman DD. Tumour necrosis factor alpha blockers are more effective than methotrexate in the inhibition of radiographic joint damage progression among patients with psoriatic arthritis. *Ann Rheum Dis* 2014;**73**:1007-11.
13. Curtis J, Yun H, FitzGerald O, Winthrop K, Azevedo V, Burmester G, et al. FRI0496 Comparing tofacitinib safety profile in patients with psoriatic arthritis in clinical studies with real-world data. *Ann Rheum Dis* 2017;**76**:676.
14. National Institute for Health and Care Excellence. *Etanercept, infliximab and adalimumab for the treatment of psoriatic arthritis*. 2010. Available from: https://www.nice.org.uk/guidance/ta199 [accessed December 2017].
15. National Institute for Health and Care Excellence. *Certolizumab pegol and secukinumab for treating active psoriatic arthritis after inadequate response to DMARDs*. 2017. Available from: https://www.nice.org.uk/guidance/ta445 [accessed December 2017].

16. 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. *Health Technol Assess* 2011;**15**:1-329.
17. *Norfolk Arthritis Register*. University of East Anglia; Available from: http://www.uea.ac.uk/noar/home [accessed January 2018].
18. Ali Y, Tom BD, Schentag CT, Farewell VT, Gladman DD. Improved survival in psoriatic arthritis with calendar time. *Arthritis Rheum* 2007;**56**:2708-14.
19. National Institute for Health and Care Excellence. *BNF*. 2017. Available from: https://bnf.nice.org.uk/ [accessed December 2017].
20. Department of Health and Social Care. *Drugs and pharmaceutical electronic market information tool (eMIT). (Last updated 5 January 2018)*. 2018. Available from: https://www.gov.uk/government/publications/drugs-and-pharmaceutical-electronic-market-information-emit [accessed March 2018].
21. National Health Service. *Reference costs 2016/17*. 2017. Available from: https://improvement.nhs.uk/resources/reference-costs/ [accessed December 2017].
22. Personal Social Services Research Unit. *Unit costs of health and social care*: Personal Social Services Research Unit; 2017.
23. National Institute for Health and Care Excellence. *Golimumab for the treatment of psoriatic arthritis*. 2011. Available from: https://www.nice.org.uk/guidance/ta220 [accessed December 2017].
24. National Institute for Health and Care Excellence. *Ustekinumab for treating active psoriatic arthritis*. 2015. Available from: https://www.nice.org.uk/guidance/ta340 [accessed December 2017].
25. Coates LC, Tillett W, Chandler D, Helliwell PS, Korendowych E, Kyle S, et al. The 2012 BSR and BHPR guideline for the treatment of psoriatic arthritis with biologics. *Rheumatology (Oxford)* 2013;**52**:1754-7.
26. National Institute for Health and Care Excellence. *Apremilast for treating active psoriatic arthritis*. 2017. Available from: https://www.nice.org.uk/guidance/ta433 [accessed December 2017].
27. Smith CH, Anstey AV, Barker JN, Burden AD, Chalmers RJ, Chandler DA, et al. British Association of Dermatologists' guidelines for biologic interventions for psoriasis 2009. *Br J Dermatol* 2009;**161**:987-1019.
28. Cummins E, Asseburg C, Punekar YS, Shore E, Morris J, Briggs A, et al. Cost-effectiveness of infliximab for the treatment of active and progressive psoriatic arthritis. *Value Health* 2011;**14**:15-23.
29. Kyle S, Chandler D, Griffiths CE, Helliwell P, Lewis J, McInnes I, et al. Guideline for anti-TNF-alpha therapy in psoriatic arthritis. *Rheumatology (Oxford)* 2005;**44**:390-7.
30. National Institute for Health and Care Excellence. *Certolizumab pegol and secukinumab for treating active psoriatic arthritis after inadequate response to DMARDs*. 2017. Available from: https://www.nice.org.uk/guidance/ta445 [accessed December 2017
31. Norton S, Fu B, Scott DL, Deighton C, Symmons DP, Wailoo AJ, et al. Health Assessment Questionnaire disability progression in early rheumatoid arthritis: systematic review and analysis of two inception cohorts. *Semin Arthritis Rheum* 2014;**44**:131-44.
32. US Food and Drug Administration. *FDA Briefing Document. Arthritis Advisory Committee Meeting, August 3, 2017*; 2017. Available from: https://www.fda.gov/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/ArthritisAdvisoryCommittee/ucm569316.htm
33. Mease PJ, Ory P, Sharp JT, Ritchlin CT, Van den Bosch F, Wellborne F, et al. Adalimumab for long-term treatment of psoriatic arthritis: 2-year data from the Adalimumab Effectiveness in Psoriatic Arthritis Trial (ADEPT). *Ann Rheum Dis* 2009;**68**:702-9.
34. Office for National Statistics. *National life tables: England and Wales*. 2017. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/lifeexpectancies/datasets/nationallifetablesenglandandwalesreferencetables [accessed January 2018].
35. Poole CD, Lebmeier M, Ara R, Rafia R, Currie CJ. Estimation of health care costs as a function of disease severity in people with psoriatic arthritis in the UK. *Rheumatology (Oxford)* 2010;**49**:1949-56.

36. Bansback NJ, Ara R, Barkham N, Brennan A, Fraser AD, Conway P, et al. Estimating the cost and health status consequences of treatment with TNF antagonists in patients with psoriatic arthritis. *Rheumatology (Oxford)* 2006;**45**:1029-38.
37. Hartman M, Prins M, Swinkels OQ, Severens JL, De Boo T, Van Der Wilt GJ, et al. Cost-effectiveness analysis of a psoriasis care instruction programme with dithranol compared with UVB phototherapy and inpatient dithranol treatment. *Br J Dermatol* 2002;**147**:538-44.

10 Appendices

Appendix A: Manufacturer's model with error in the implementation of the placebo-response adjustment

Corrected placebo-response adjustment model:

```
model{
for(i in 1:NS){
            mu[i] ~ dnorm(0,.0001)
            w[i,1] <-0
            delta[i,1]<-0      for (k in 1:na[i])  {
                  r[i,k] ~ dbin(p[i,k],n[i,k])
                  logit(p[i,k])<-mu[i] + delta[i,k]  + (beta[t[i,k]]-beta[t[i,1]]) *(mu[i]-meanmA)
                  }
          for (k in 2:na[i]) {
            delta[i,k] ~ dnorm(md[i,k],taud[i,k])
            md[i,k] <-  d[t[i,k]] - d[t[i,1]]  + sw[i,k]
             taud[i,k] <- tau *2*(k-1)/k
             w[i,k] <- (delta[i,k]  - d[t[i,k]] + d[t[i,1]])
            sw[i,k] <-sum(w[i,1:k-1])/(k-1) }
          }
          d[1]<-0
          beta[1] <- 0
          for (k in 2:NT){
                  d[k] ~ dnorm(0,.0001)
                  beta[k] <- B
```

```
            }
        B ~ dnorm(0,.0001)
        sd~dunif(0.001,2)
        tau<-1/pow(sd,2)
}}
```

Appendix B: Comparison of costs and QALYs between TA445 and TA1220

Treatment	Average QALYs in TA445	QALYs in TA1220	Average costs in TA445	Costs in TA1220
Sub-population 2				
BSC	5.7	[redacted]	£71467	[redacted]
ADA	7.7	[redacted]	£117680	[redacted]
CTZ	7.5	[redacted]	£115719	[redacted]
ETN	8.1	[redacted]	£123167	[redacted]
SEK	7.6	[redacted]	£120409	[redacted]
GOL	8.0	[redacted]	£123123	[redacted]
INF	8.2	[redacted]	£148786	[redacted]
Sub-population 3				
BSC	5.7	[redacted]	£71467	[redacted]
UST	6.7	[redacted]	£95362	[redacted]
SEC	7.0	[redacted]	£122357	[redacted]
Sub-population 4				
BSC	5.7	[redacted]	£71467	[redacted]
UST	6.7	[redacted]	£92404	[redacted]
SEC	6.8	[redacted]	£99764	[redacted]