NICE National Institute for Health and Care Excellence

Single Technology Appraisal

Certolizumab pegol for treating rheumatoid arthritis after inadequate response to a TNF inhibitor

Committee Papers

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

SINGLE TECHNOLOGY APPRAISAL

Certolizumab pegol for treating rheumatoid arthritis after inadequate response to a TNF inhibitor [ID824]

Contents:

Any information supplied to NICE which has been marked as confidential, has been redacted. All personal information has also been redacted.

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

Abbreviated premeeting briefing

Certolizumab pegol for treating rheumatoid arthritis after inadequate response to a TNF Inhibitor [ID824]

This abbreviated premeeting briefing highlights key issues for discussion at the first Appraisal Committee meeting and should be read with the full supporting documents for this appraisal.

Please note that this document includes information from the ERG before the Company has checked the ERG report for factual inaccuracies.

1 Technology

Technology	Certolizumab pegol solution for injection (Cimzia, UCB Pharma)
Class of drug	Certolizumab pegol is a recombinant, humanised antibody Fab' fragment against tumour necrosis factor alpha (TNFα), a pro-inflammatory mediator that is partly responsible for damage to the joints in rheumatoid arthritis, expressed in *Escherichia coli* and conjugated to polyethylene glycol (PEG)
Administration method	Subcutaneous injection
List price	200-mg prefilled syringe = £357.50 (BNF)
Patient access scheme (PAS)	A PAS is available to the NHS where the first 12 weeks of treatment are free (10 vials) = £3575

Annual cost of treatment (including PAS)	Loading doses at week 0, 2 and 4 of 400mg and maintenance doses of 200mg every 2 weeks or 400mg every 4 weeks once clinical response is confirmed = (first year) £6793 (PAS) and £9295 (after first year PAS)
Marketing authorisation	Marketing authorisation in the UK for the treatment of moderate to severe, active rheumatoid arthritis in adult patients when the response to disease modifying anti-rheumatic drugs (DMARDs), including methotrexate, has been inadequate. Certolizumab pegol can be used in combination with methotrexate, or as a monotherapy if methotrexate is not tolerated or is contra-indicated. It can also be used for the treatment of severe, active and progressive RA in adults not previously treated with methotrexate or other DMARDs.
SmPC	Link to report
EPAR	Link to report
Abbreviations: BNF; British National Formulary, EPAR; European public assessment report, mg; milligram, PAS; patient access scheme, SMPC; summary of product characteristics, TA; technology appraisal	

2 Relevant appraisals and current treatment pathway

The Committee has considered biological disease modifying anti-rheumatic drugs (bDMARDs) in previous technology appraisals (TAs), the two main being multiple technology appraisals (MTAs) TA375 and TA195 (Table 1). TA375 considers the use of bDMARDS at an earlier position in the treatment pathway (i.e. after the failure of conventional DMARDs (cDMARDs) than TA195 which considers their use after the failure of a TNF inhibitor (TNFi). Certolizumab pegol was not considered within TA195 as it did not have a marketing authorisation at the time of the appraisal; it is therefore considered separately here. The recommendations from these two MTAs set out a

treatment pathway in which a combination of intensive cDMARDs (including methotrexate) should be the first treatment option, and if the disease fails to respond, then one of the following drugs may be used: adalimumab, etanercept, infliximab, certolizumab pegol, golimumab, tocilizumab and abatacept (see table 1 for full details of the recommendation from TA375). Of these, adalimumab, etanercept, golimumab, infliximab and certolizumab pegol are TNF inhibitors. For adults who have had an inadequate response to, or are intolerant of other DMARDs, including at least one TNF inhibitor, rituximab (in combination with methotrexate) should then be tried (see table 1 for full details of the recommendation from TA195). The recommendations provide alternative options when either methotrexate or rituximab are contraindicated or not tolerated. In this appraisal, the Committee is asked to consider whether certolizumab pegol (in combination with methotrexate or as a monotherapy) should be recommended where the existing bDMARDs have been recommended in TA195 and in addition, the Company would like the Committee to consider whether it can be recommended ahead of rituximab, following the failure of at least one TNF inhibitor.

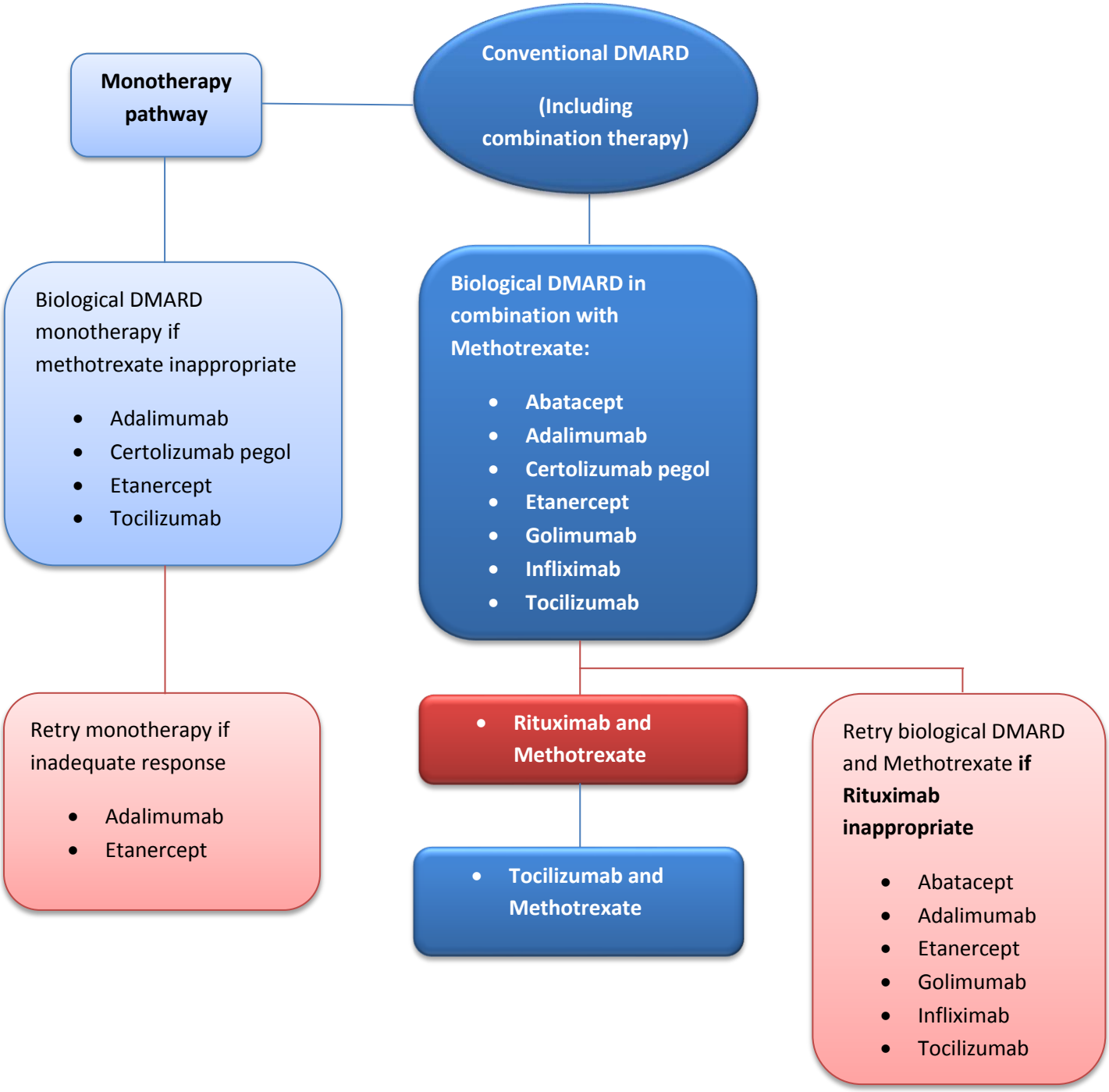
Table 1 Current bDMARD recommendations

TA375 Adalimumab, etanercept, infliximab, certolizumab pegol, golimumab, tocilizumab and abatacept for treating rheumatoid arthritis not previously treated with DMARDs or after conventional DMARDs only have failed	**TA195** Adalimumab, etanercept, infliximab, rituximab and abatacept for treating rheumatoid arthritis after the failure of a TNF inhibitor
Recommended: • Disease is severe, that is, a disease activity score (DAS28) greater than 5.1 • Disease has not responded to intensive therapy with a combination of conventional disease-modifying drugs (DMARDs) and • The companies provide certolizumab pegol, golimumab, abatacept and	Recommended: • Rituximab in combination with methotrexate is an option for severe RA in adults who have had inadequate response to, or are intolerant of, other DMARDs, including at least one TNFi. • Adalimumab, etanercept, infliximab and abatacept each in combination with methotrexate in severe RA in adults who

tocilzumab as agreed in their patient access scheme. • Certolizumab pegol, etanercept, adalimumab or tocilizumab can also be used as monotherapy when methotrexate is contra-indicated or intolerant	have an inadequate response to, or intolerant to, other DMARDs, including one other TNFi and who cannot receive rituximab because they have a contra-indication to methotrexate or is withdrawn to an adverse event. • Adalimumab and etanercept monotherapy in severe RA in adults who have had an inadequate response or have an intolerance to, other DMARDs, including one other TNFi, and who cannot receive rituximab because methotrexate is contra-indicated or withdrawn due to adverse events.

Certolizumab pegol (also referred to as certolizumab) has not been assessed in people who have had inadequate response to other first line TNF inhibitor. The Company cite TA375 as a basis that certolizumab was shown to be as effective other available TNFi and an alignment with bDMARDs after initial failure with a TNF inhibitor is needed, which is evident in European League Against Rheumatism (EULAR) guidelines.

Figure 1 Summarised NICE treatment pathway for RA. Red pathways show Company's proposed Certolizumab pegol placement



3 Decision problem

Table 2 PICO table from PICO table from the NICE scope (including indication of adherence/deviations in Company's submission)

	Decision problem addressed in the submission	✓ / ✗
Population	Adults with moderate to severe active rheumatoid arthritis, defined as disease activity score 28 (DAS28)>3, whose disease has not responded adequately to a tumour necrosis factor (TNF) inhibitor (TNFi). This population selection was based on the British Society for Rheumatology (BSR) recommendations for treatment with TNFi.	✓
Intervention	CIMZIA® (certolizumab pegol, CZP) monotherapy or in combination with methotrexate (MTX).	✓
Comparators	For adults previously treated with other disease-modifying antirheumatic drugs (DMARDs) including at least 1 TNFi • Rituximab (RTX) in combination with methotrexate (MTX)	✓
	For adults for whom RTX is contraindicated or withdrawn • abatacept (ABA), adalimumab (ADA), etanercept (ETA), golimumab (GOL), infliximab (IFX) and tocilzumab (TOC) each in combination with MTX	✓
	For adults for whom RTX therapy cannot be given because MTX is contraindicated or withdrawn. • ADA monotherapy, ETA monotherapy or TOC monotherapy	✓
	• Best supportive care	✗
Outcomes	Manufacturer included patient reported outcomes (PROs) and other outcome measures which includes: • Disease activity	✓
	• Physical function	✓
	• Joint damage	✗
	• Pain	✓
	• Mortality	✓
	• Fatigue	✓

	• Radiological progression • Extra-articular manifestations of the disease • Adverse effects of treatment • Health-related quality of life	✗ ✗ ✓ ✓

4 Summary of clinical evidence

The Company's systematic literature review identified six randomized controlled trials and state that the 'largest breadth' of evidence comes from the REALISTIC trial, which included the largest number of relevant patients with prior TNFi use (TNFi-experienced) whereas J-RAPID and HIKARI had low numbers of patients with previous experience with a TNFi. The primary outcome from five of the trials was an American College of Rheumatology improvement score of 20% (ACR20, results also included improvements at 50% and 70%), whereas in the PREDICT trial the focus was on CDAI and RAPID-3 assessment scores and disease-activity score (DAS28(ESR) in 28 joints assessed). EULAR responses were shown (none, moderate or good) as a secondary outcome (not reported from SWITCH). Quality of life data was collected in the form of the short-form 36 from DOSEFLEX and EuroQol data from PREDICT (EQ-5D – but results not included in Company's clinical-effectiveness analysis). Other patient reported outcomes were collected in the form of the quality of life health assessment questionnaire disability index (HAQ-DI, lower scores indicating improvement) and the impact on fatigue and sleep (see section 4.3, page 49 for full details of study design).

Table 3 List of relevant RCTs in the Company's submission

Trial	Study Design	Interventions	Patient population	Duration of study	Primary outcome(s)
REALISTIC (n=1,063, TNFi-IR= 400 conducted in USA, Canada, France, Germany, The Netherlands and Spain) (NCT00717236)	• Phase III randomised double-blind, placebo controlled for 12 weeks, with an open-label extension for up to at least week 16	• CZP 200 mg Q2W +/- MTX/cDMARDs* • PBO +/- MTX/cDMARDs	Active RA with an inadequate response to >1 prior cDMARD, having received treatment with ≤ 2 TNFis	28 weeks	• ACR20 response at 12 weeks
DOSEFLEX (n= 333, TNFi-IR= 178 conducted in USA, Canada and France) (NCT00580840)	• Phase IIIb multicentre, double-blind randomisation, open-label run-in for 16 weeks, then placebo-controlled up to week 34	• CZP 200 mg Q2W + MTX[≠] • CZP 400 mg Q4W + MTX[≠] • PBO + MTX[≠]	Active RA receiving MTX for ≥3 months, including patients with prior TNFi exposure	34 weeks[≠]	• ACR20 response at 34 weeks
PREDICT (n=733, TNFi-IR=407 conducted in USA) (NCT01255761)	• Phase IV, multicentre, double-blind, randomisation, placebo-controlled up to 52 weeks	• CZP 200 mg Q2W +/- MTX/cDMARDs*	Active RA with unsatisfactory response or intolerance to ≥1 DMARD, having received treatment with ≤2 TNFis	52 weeks	• CDAI and RAPID-3 scores at 12 and 52 weeks • DAS28(ESR) at 52 weeks
SWITCH (Only TNFi-IR= 37 conducted in USA) (NCT01147341)	• Phase IV, multicentre, double-blind, randomisation, placebo-controlled up to 12 weeks	• CZP 200 mg Q2W + cDMARDs* • PBO + cDMARDs	Active RA having had inadequate response or intolerance to a TNFi other than CZP	24 weeks	• ACR20 response at 12 weeks
J-RAPID[§] (n= 159, TNFi-IR= 26 conducted in Japan) (NCT00791999)	• Phase II/III multicentre, double-blind, randomisation, placebo-controlled up to week 16 then open-label extension up to week 24	• CZP 100 mg Q2W + MTX ** • CZP 200 mg Q2W + MTX * • CZP 400 mg Q2W + MTX * • PBO + MTX	Active RA with an inadequate response to MTX, including patients with prior exposure if they received 1 TNFi as a non-primary failure (only Japanese patients)	24 weeks	• ACR20 response at 12 weeks
HIKARI (n= 230, TNFi-IR=16 conducted in Japan) (NCT00791921)	• Phase III double-blind, randomisation, placebo-controlled up to week 16 then open-label extension up to week 24	• CZP 200 mg Q2W +/- non-MTX cDMARDs[∓] • PBO +/- non-MTX cDMARDs	Active RA with an inadequate response to ≥1 prior DMARDs (including MTX), including patients with prior exposure if they received 1 TNFi as a non-primary failure (only Japanese patients)	24 weeks	• ACR20 response at 12 weeks

*Patients received a loading dose of 400 mg CZP at Weeks 0, 2, and 4; **Patients received a loading dose of 200 mg CZP at Weeks 0, 2, and 4,; ≠ In DOSEFLEX, all patients received 400 mg CZP at Weeks 0, 2, and 4 followed by 200 mg CZP Q2W up to and including Week 16 in a run-in phase, followed by 16 weeks in randomisation to listed interventions.

§Trial contains CZP doses unlicensed in the European Union (CZP loading dose of 200 mg at Weeks 0, 2 and 4, for the CZP 100 mg Q2W dosage arm, and maintenance doses of CZP 100 mg Q2W and CZP 400 mg Q2W). ₸CZP in combination with non-MTX cDMARDs is not approved in the European Union;

TNFi-IR; Patients with previous exposure to a TNFi , Q2W; every 2 weeks, PBO; placebo, n; number of a type of patient in the trial

Source Adapted from Company's submission Table 9

Clinical trial Results

ACR Response rates

In all six trials the ACR response was higher in the certolizumab group compared with placebo for the primary outcome at the primary time point. In REALISTIC, DOSEFLEX, PREDICT and HIKARI, they were comparable to the overall population. The Company's graphs show that all responses are higher at the measured time points (See Company's submission Section 4.7, page 77 for full trial results).

Table 4 ACR response rates in the TNFi-experienced populations

Trial	Treatment Group	Treatment arms for which data extraction performed (n)	Assessment time point	% achieving ACR20 response
REALISTIC	TNFi-experienced	[redacted]	[redacted]	[redacted]
		[redacted]	[redacted]	[redacted]
	TNFi-experienced (NRI), CZP monotherapy	[redacted]	[redacted]	[redacted]
		[redacted]	[redacted]	[redacted]
		[redacted]	[redacted]	[redacted]
		[redacted]	[redacted]	[redacted]
	TNFi-experienced (NRI), CZP+MTX	[redacted]	[redacted]	[redacted]
		[redacted]	[redacted]	[redacted]
		[redacted]	[redacted]	[redacted]

Trial	Treatment Group	Treatment arms for which data extraction performed (n)	Assessment time point	% achieving ACR20 response
		[redacted]	[redacted]	[redacted]
DOSEFLEX	TNFi-experienced (NRI)	[redacted]	[redacted]	[redacted]
	TNFi-experienced (LOCF)	[redacted]	[redacted]	[redacted]
SWTICH	TNFi-experienced	PBO Q2W + cDMARDS (n=10)	Week 12	0 (0%)
		CZP 200 mg Q2W+cDMARDS (n=27)	Week 12	17 (61.5%)
		PBO Q2W + cDMARDS (n=8)	Week 24	5 (62.5%) (P value NR)
		CZP 200 mg Q2W+cDMARDS (n=22)	Week 24	12 (54.5%) (P value NR)
HIKARI	TNFi-experienced	PBO (n=10)	Week 12	[redacted]
		CZP 200 mg Q2W (n=6)	Week 12	[redacted]
		PBO (n=10)	Week 24	[redacted]
		CZP 200 mg Q2W (n=6)	Week 24	[redacted]
J-RAPID	TNFi-experienced	PBO +MTX (n=15)	Week 12	[redacted]
		CZP 200 mg Q2W +MTX (n=11)	Week 12	[redacted]
		PBO +MTX (n=15)	Week 24	[redacted]
		CZP 200 mg Q2W +MTX (n=11)	Week 24	[redacted]

LOCF; last observation carried forward, NRI; non-responder imputation (both statistical analytical methods for the intention-to-treat population), +; indicates combination treatment with.., NR; not reported, P; statistical significance value. Source Adapted form ERG report Table 12

EULAR response

The Company's figures showed that the number of responses for patients in the [redacted]. Of note; the REALISTIC study showed that the response rate was higher for the certolizumab and methotrexate combination group over the monotherapy [redacted] The ERG provided results for the SWTICH trial from ClinicalTrials.gov and showed that certolizumab treated patients demonstrated a much more favourable EULAR response at week 12 and results at week 24 were roughly comparable between certolizumab and placebo groups (no p-values reported; see ERG report section 4.2.6, page 70 for a detailed summary of the Company's EULAR response results).

Meta-Analysis

A direct meta-analysis was undertaken in order to pool the data from the REALISTIC, J-RAPID and SWITCH trials to compare the sub-population of patients for certolizumab in combination with methotrexate and methotrexate in the placebo arms. The Company did not include DOSEFLEX or PREDICT in the meta-analysis. A direct meta-analysis was also conducted to pool data from REALISTIC and HIKARI for the sub-populations of patients that receive cetrolizumab monotherapy verses placebo. Statistical heterogeneity was

detected (using Higgins I^2) therefore both fixed and random effects models were used and presented (see Company's submission section 4.9, page 117).

Table 5 Results of direct meta-analysis for certolizumab (combination with methotrexate) vs methotrexate and for certolizumab (monotherapy) vs placebo, at 3 months (week 12). Risk ratios above 1 favour the intervention

	ACR20 response at 3 months RR (95% CI)	**ACR50 response at 3 months RR (95% CI)**	**ACR70 response at 3 months RR (95% CI)**	**EULAR (good)**	**EULAR (good to moderate)**
Fixed effect model **(Combination)**	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Random effects model **(Combination)**	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Fixed effect model **(Monotherapy)**	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Random effects model **(Monotherapy)**	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
ACR, American College of Rheumatology criteria; CI, confidence interval; RR, relative risk.					

Source Adapted from Company's appendices to the submission Table 8.11.1 and 8.11.2.

**Note; miscalculated upper CI from Company tables.*

Network Meta-Analysis and adjusted indirect comparisons; certolizumab compared with other bDMARDs

The Company identified nine RCTs for an indirect and mixed treatment comparison using an adjusted indirect comparison (ITC) method and Bayesian network meta-analysis (NMA) to assess the comparative effectiveness for certolizumab with other comparator bDMARDs. Of note; the network of diagrams from the Company show that the largest trial, REALISTIC, was only included for comparisons at the 3 month time points whereas comparisons made at 6 months included the smaller J-RAPID and

other trials (please refer to Company's submission section 4.10, page 123 for a detailed account of the methods used and Company's appendices section 8.12.3, page 42 for the network of evidence diagrams used in the NMA and ITC).

The Company's base case results show that certolizumab in combination with methotrexate is at least as effective to the other comparators in patients with moderate to severe RA with pervious exposure, noting that the 'wide credible' confidence intervals show minimal differences in the clinical effect between certolizumab and all comparators but results should be interpreted with caution due heterogeneity among patient population.

The Company's results show that there was no significant difference observed when certolizumab in combination with methotrexate was compared with tocilizumab and methotrexate, and abatacept with methotrexate for an ACR20 response rate at 3 months and no difference when compared to tocilizumab with methotrexate, for an ACR50 response rate at 3 months. Results were in favour of certolizumab and methotrexate when compared to abatacept, golimumab, rituximab and tocilizumab for an ACR20, 50 and 70 response at 6 months but not statistically significant. No significant differences were observed between certolizumab and rituximab in combination with methotrexate for a EULAR good/moderate response at 3 months ([redacted]) Please see section 4.10.6, page 128 of Company's submission for full NMA and ITC results)

Figure 2 Forest plot for ITC for ACR20 at 3 months; showing relative risks (RRs) with 95% confidence intervals (CI)

Source Company's submission Figure 44

Figure 3 Forest plot for ITC for EULAR (good/moderate) response at 3 months: showing RRs with 95% Cis

Source Company's submission Figure 50

ERG comments on clinical effectiveness

The ERG comment that all six trials were of good quality and the literature search strategy was logical and consistent but says that one certolizumab RCT was left out of the analysis (Kang et al. 2012) due to low TNFi-experienced patients, yet included two other trials with low numbers (J-RAPID and HIKARI). The ERG said that a more robust justification is needed for excluding Kang et al. Often statistical significance of results was not included (p-values) and that wide credible interval (confidence intervals) should be interpreted with cation (true effect uncertain and does not show minimal difference between results; see ERG report section 4.5, page 109 for full details).

5 Summary of cost effectiveness evidence

The Company submitted a cohort Markov model in which certolizumab could be evaluated as a treatment for 3 possible populations:

- Patients for whom rituximab plus methotrexate is a treatment option (referred to as population A)
- Patients for whom rituximab is contraindicated or withdrawn (referred to as population B)
- Patients for whom methotrexate is contraindicated or withdrawn due to an adverse event (referred to as population C)

Table 6 Population A: Treatment sequence considered by the Company

Intervention Sequence	Comparator sequence
CZP + MTX	RTX + MTX
RTX + MTX	TOC + MTX
TOC + MTX	ABA+MTX
ABA+MTX	MTX + hydroxychloroquine + sulfasalazine

Intervention Sequence	Comparator sequence
MTX + hydroxychloroquine + sulfasalazine	Non-biologic (Weighted mix of leflunomide, gold, ciclosporin, azathioprine (25% each))
Non-biologic (Weighted mix of leflunomide, gold, ciclosporin, azathioprine (25% each))	Palliative care
Palliative care	-

Source Company's submission Table 64

Table 7 Population B: Treatment sequence considered by the Company

Intervention Sequence (	Comparator biologic + MTX
Certolizumab + methotrexate	Comparator bDMARD + methotrexate
MTX + hydroxychloroquine + sulfasalazine	MTX + hydroxychloroquine + sulfasalazine
Leflunomide	Leflunomide
Gold injection	Gold injection
Ciclosporin	Ciclosporin
Azathioprine	Azathioprine
Palliative care	Palliative care

Source Company's submission Table 65

Table 8 Population C: Treatment sequence considered by the Company

Intervention sequence	Comparator sequences
CZP	Comparator biologic
Leflunomide	Leflunomide
Gold injection	Gold injection
Ciclosporin	Ciclosporin
Azathioprine	Azathioprine
Palliative care	Palliative care

Source Company's submission Table 66

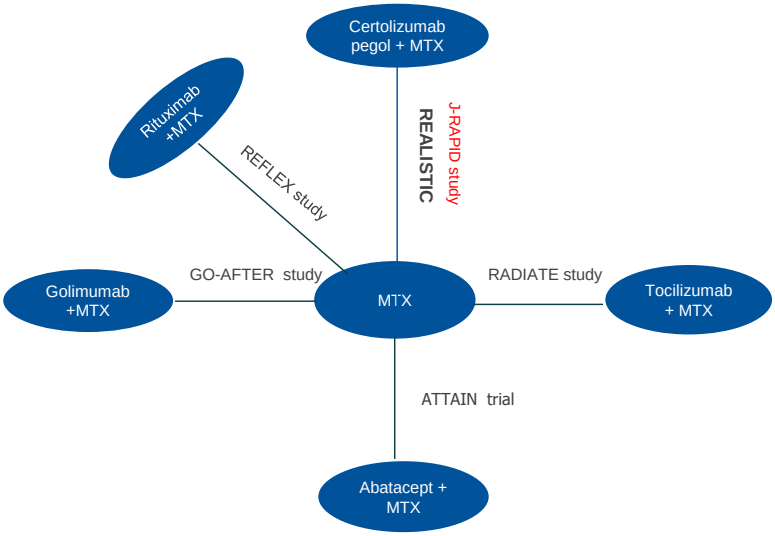
The baseline characteristics of all 3 of the modelled population were based on mean estimates from the REALISTIC trial ([redacted]

[redacted] (see section 5.2.1, page 166 of the Company's submission)

The model schematic can be found on page 172 (figure 57) of the Company's submission with health states based on patients EULAR response criteria (no response, moderate or good) to the first treatment. This assumes that patients entering the model, after initial failure with TNFi, start with a bDMARD immediately. There are 2 states for each of the seven subsequent treatments (one state to represent the first six months of treatment, and one state for the remainder of time on that treatment); and death states. Patients assigned to a non-responder state discontinue their first therapy and start the next one in the sequence (see section 5.2.5 of Company submission for details of comparator sequences in the specific populations modelled). The model has a 45 year time horizon and costs and benefits are discounted at 3.5% with half-cycle corrections and the model takes a NHS/PSS perspective (see section 5.2.2 and 5.2.3, page 167 of Company submission for full details of model structure and clinical pathway).

Clinical response to first treatment was model via the EULAR response probabilities. These were estimated from effect size estimates and cut-off probabilities, taken from a Bayesian network meta-analysis from the previous analysis (see NMA and ITC above and section 4.10.6, page 128 of the Company's submission) which imposed a series of assumptions on comparative effectiveness between biologics. The Company chose the REALISTIC trial for the baseline characteristics and response probabilities for EULAR no response rates. The Company included the J-RAPID trial in a sensitivity analysis (network diagram shown below) and [redacted]. The Company choose a fixed-effect model for the NMA and results for the EUALAR probabilities are shown below (see section 5.3, page 176 of the Company submission for full details)

Figure 4 Network of evidence considered in the Company's base case analysis (J-RAPID included in Company sensitivity analysis and ERG base case)



Source Company's submission Figure 62

Figure 5 Estimated Mean EULAR response probabilities from the NMA Population A

Source Company's submission Figure 58

Figure 6 Estimated Mean EULAR response probabilities from the NMA Population B

Source Company's submission Figure 59

Figure 7 Estimated Mean EULAR response probabilities from the NMA for Population C

Source Company's submission Figure 60

Utilities were modelled splitting health effects into two elements. Relating initial response to EQ5D (via a linear regression model from patient-level data

form the PREDICT trial) for the first 6 months and mapping from HAQ score to EQ-5D following 6 months (see from section 5.3.2, page 193 to 211 for full details on how the Company modelled the various assumptions). Adverse effects were not incorporated in the Company's model. Costs included various assumptions for weight distribution, vial use and dosing (see Company's submission page 215).

Company base case results

The Company presented an incremental cost effectiveness analysis incorporating the agreed complex schemes for certolizumab and golimumab. Confidential PAS discounts also exist for abatacept and tocilizumab but are not included in the company base case results. Results including these discounts have been provided by the ERG in a confidential appendix for the committee.

Table 9 Population A; Base case results for patients eligible for rituximab and methotrexate

Sequences	Total QALYs	Total costs	Inc. QALYs	Inc. costs	ICER (£/QALY)	Probability (%) of cost effectiveness at a threshold of	
						£20,000/ QALY	£30,000/ QALY
RTX‡ **(Deterministic)**	7.000	£138,520	-	-	**-**		
CZP before RTX† **(Deterministic)**	7.286	£148,361	0.286	£9,842	**£34,378**		
RTX † (**Probabilistic)**	7.031	£139,933	-	-	**-**	97.80	63.02
CZP before RTX† **(Probabilistic)**	7.321	£149,579	0.290	£9,647	**£33,222**	2.20	36.98

Treatment sequence: ‡ RTX: RTX+MTX, TOC+MTX, ABA+MTX, MTX + HCQ (hydroxychloroquine) + SSZ (sulfasalazine), NBT (non-biological therapy), PC (palliative care) † CZP before RTX: CZP+MTX, RTX+MTX, TOC+MTX, ABA+MTX, MTX + HCQ + SSZ, NBT, PC. *Source Adapted from ERG report Table 40 and 41*

Table 10 Population B; Base case results for patients whom rituximab is inappropriate

First therapy of the sequence†	Total QALYs	Total costs	Inc. QALYs	Inc. costs (£)	ICER (£/QALY)	Probability(%) of cost effectiveness at a threshold of	
						£20,000/ QALY	£30,000/ QALY
IFX + MTX **(Deterministic)**	6.048	£101,484	-	-	**Dominated**	-	-
ETA + MTX **(Deterministic)**	6.048	£97,606	-	-	**Dominated**	-	-
ADA + MTX **(Deterministic)**	6.048	£97,183	-	-	**-**	-	-
GOL + MTX **(Deterministic)**	6.048	£97,183	-	-	**-**	-	-
ABA(IV) + MTX‡ **(Deterministic)**	6.095	£115,555	0.047	£18,373	**Dominated**	-	-
CZP + MTX **(Deterministic)**	6.308	£98,100	0.260	£918	**£3,527**	**-**	**-**
TOC(IV) + MTX‡ **(Deterministic)**	6.507	£125,112	0.199	£27,011	**£135,953**	**-**	**-**
IFX + MTX **(Probabilistic)**	6.038	£102,242	-	-	**Dominated**	0.00	0.00
ETA + MTX **(Probabilistic)**	6.070	£98,360	-	-	**Dominated**	0.0	0.7
GOL + MTX **(Probabilistic)**	6.071	£97,964	-	-	**-**	0.3	1.5
ADA + MTX **(Probabilistic)**	6.076	£98,015	-	-	**Extendedly dominated**	0.2	1.7
ABA (IV)+ MTX‡ **(Probabilistic)**	6.119	£116,232	-	-	**Dominated**	0.00	0.00
CZP + MTX **(Probabilistic)**	6.327	£98,848	0.256	£884	**£3,461**	99.5	96.0
TOC (IV)+ MTX‡ **(Probabilistic)**	6.528	£125,507	0.201	£26,659	**£132,783**	0.00	0.00

†Rest of the sequence: MTX + HCQ + SSZ, LEF (leflunomide), GLD (gold), CIC (ciclosporin), AZA (azathioprine), PC; ‡CiC PAS not included. *Source Adapted form ERG report Table 42 and 43*

Table 11 Population C Base case results for patients for whom methotrexate is inappropriate

First therapy of the sequence†	Total QALYs	Total costs	Inc. QALYs	Inc. costs	ICER (£/QALY)	Probability (%) of cost effectiveness at a threshold of	
						£20,000/ QALY	£30,000/ QALY
ADA **(Deterministic)**	5.880	£95,632	-	-	-		
ETA	5.880	£96,036	-	-	Dominated		

(Deterministic)							
CZP **(Deterministic)**	6.141	£97,249	0.260	£1,617	**£6,213**		
TOC (IV) ‡ **(Deterministic)**	6.346	£123,592	0.206	£27,960	**£127,955**		
ETA **(Probabilistic)**	5.899	£96,270	-	-	**Dominated**	0.04	0.92
ADA **(Probabilistic)**	5.902	£95,918	-	-	**-**	0.18	1.16
CZP **(Probabilistic)**	6.162	£97,254	0.260	£1,336	**£5,151**	99.78	97.48
TOC(IV) ‡ **(Probabilistic)**	6.358	£123,433	0.196	£26,179	**£133,655**	0.00	0.00

† Rest of the sequence: LEF, GLD, CIC, AZA, PC; ‡ CiC PAS not included. *Source Adapted form ERG report Table 44 and 45*

Sensitivity Analysis

The Company performed a one-way and scenario analyses to test the robustness of results. The parameters exhibiting most sensitivity were changes to varying the treatment effect on probability of EULAR response to the extremes of the 95% CI for certolizumab and its comparators, the discount rates and altering the mapping algorithm for HAQ to EQ-5D. All these showed wide percentage changes from the base case net monetary benefit values. Population A and B were most sensitive to changes to treatment effect of certolizumab and its comparators, while population C was most sensitive to changes in the discount rates for cost and health effect (see section 5.8.2, page 241 of the Company submission for full details and forest plots).

In the Company structural exploration (scenario analysis) of the model, all populations were most sensitive to changes in the time-horizons, applying a societal perspective and the duration of bDMARDs after response. Population A was also sensitive to the mapping algorithm applied from HAQ to EQ-5D and the efficacy of certolizumab compared to placebo. In further analysis the Company included the sub-cutaneous (SC) formulations of tocilizumab and abatacept and biosimilars for infliximab (inflectra and remsima) (see section 5.8.3, page 245 for full details).

Table 12 scenario analysis conducted by the Company

Parameter	Base case	Scenario	Certolizumab pegol ICER (£/QALY)		
			Population A	**Population B**	**Population C**
Efficacy of CZP	Based on NMA	Pop. A and B = other TNFi and pop C = ADA and ETA	**£169,690**	-	-
		Including J-RAPID in NMA	**£29,613**	**£4,000**	**£7,000**
Source of utility for first treatment response	Linear regression using PREDICT	HAQ scores from REALISTIC mapped to EQ-5D	**£33,199**	**£6,000**	**£8,000**
RTX retreatment interval	6 months	9 months	**£49,618**	**-**	**-**
SC formulation of TOC and ABA and IFX biosimilar in pop. B	IV	SC	**-**	**£3,641**	**-**
SC formulation of TOC in pop. C	IV	SC	**-**	**-**	**£4,985**

Source Adapted from ERG report Table 49

ERG Critique of cost-effectiveness and preferred base case

The ERG provided a number of comments relating to the modelling approach used by the Company as it is not the most appropriate to reflect the nature of the disease and that an individual patient-level model would be preferable that can capture non-linear functions and track time in treatment. It says that cohort models cannot, as admitted by the Company, accurately reflect non-linear functions such as mapping of HAQ scores to the EQ-5D preferred by the ERG (see section 5.3, page 148 of the ERG report). The ERG note that EQ-5D data used was from the PREDICT study but the baseline population characteristics for the model were from the REALISTIC study therefore the utilities used in the model are biased (see page 148 of ERG report). Regarding the implementation of the model, the ERG carried out model validation and corrected some technical programming errors (see section 5.4 page 149 of the ERG report). The ERG provided an exploratory preferred base case ICER in which a number of amendments were made:

- Correction of technical programming errors in the Company's model.
- Adding two other sequences to be compared for Population A.
- Removing abatacept treatment from the intervention and comparator sequences for Population A.
- Using the results of the NMA including J-RAPID.
- Setting rituximab retreatment interval to 7.35. The Appraisal Committee for TA195 concluded that the average retreatment interval was between 6 and 8.7 months. The ERG used the midpoint between these two figures: (6+8.7)/2= 7.35.
- Using different HAQ improvement for subsequent therapies. Instead of the -0.39 and -0.05 mean change in HAQ score for responders to subsequent bDMARD and cDMARD treatments respectively. Values of -0.576 for bDMARD responders and -0.303 for cDMARD responders were used instead.
- Using the Weibull parameters reported in TA195 for rituximab (see **Error! Reference source not found.** of the ERG report) instead of assuming the same time to discontinuation as for TNF inhibitors.

- Assume that mortality is only affected by the baseline HAQ score, and that changes in the HAQ score do not affect mortality.
- Using constant discontinuation rates for subsequent bDMARD treatments that would match the mean treatment duration estimated by the Weibull distribution used for the first treatment line considered in the model (see **Error! Reference source not found.** of ERG report).
- Including the 80 mg dose of TOC (IV) and 800 mg limit for people with a body weight greater than 100 kg.
- Using amended numbers of administrations per cycle for infliximab (3.25) and tocilizumab IV (7 in the first cycle).
- Including the sub-cutaneous formulations of abatacept and tocilizumab, infliximab biosimilars and Benepali (a new ETA biosimilar) as comparators in its analyses.

Table 13 ERG preferred base case results for population A

Sequences	Total QALYs	Total costs	Inc. QALYs	Inc. costs	ICER (£/QALY)	Probability (%) of cost-effectiveness at a threshold of	
						£20,000/ QALY	£30,000/ QALY
CZP instead of RTX‡ **(Deterministic)**	7.719	£125,364	-	-	**Dominated**	-	-
CZP before RTX‡ **(Deterministic)**	8.239	£133,780	-	-	**Dominated**	-	-
RTX‡ **(Deterministic)**	8.378	£122,451	-	-	**-**	-	-
CZP after RTX‡ **(Deterministic)**	8.649	£130,016	0.271	£7,565	**£27,946**	**-**	**-**
CZP instead of RTX‡ **(Probabilistic)**	7.796	£128,376	-	-	**Dominated**	0.00	0.00
CZP before RTX‡ **(Probabilistic)**	8.347	£136,751	-	-	**Dominated**	0.00	0.20
RTX‡ **(Probabilistic)**	8.461	£125,189	-	-	**-**	71.46	45.64
CZP after RTX‡ **(Probabilistic)**	8.732	£132,692	0.271	£7,504	**£27,700**	28.52	54.26

†Rest of the sequence: TOC(SC)+MTX, MTX + HCQ + SSZ, NBT, PC
‡CiC PAS not included; *Source Adapted from ERG report Table 58 and 59*

Table 14 ERG preferred base case results for population B; only showing those therapies that have not been dominated.

First therapy of sequence	Total QALYs	Total costs	Inc. QALYs	Inc. costs	ICER (£/QALY)	Probability (%) of cost-effectiveness at a threshold of	
						£20,000/ QALY	£30,000/ QALY
CZP + MTX‡ **(Deterministic)**	7.176	£95,197	0.279	£3,562	**£12,773**	-	-
TOC(SC) + MTX ‡ **(Deterministic)**	7.697	£118,338	0.520	£23,141	**£44,479**	-	-
CZP + MTX **(Probabilistic)**	7.213	£95,899	0.280	£3,392	**£12,116**	96.22	92.30
TOC(SC) + MTX **(Probabilistic)**	7.725	£119,171	0.571	£23,272	**£45,414**	0.000	0.0028

†Rest of the sequence: MTX + HCQ + SSZ, LEF, GLD, CIC, AZA, PC; ‡CiC PAS not included; bio = biosimilar; *Source Adapted from ERG report Table 60 and 61*

Table 15 ERG preferred base case results for population C; only showing those therapies that have not been dominated.

First therapy of sequence	Total QALYs	Total costs	Inc. QALYs	Inc. costs	ICER (£/QALY)	Probability (%) of cost-effectiveness at a threshold of	
						£20,000 / QALY	£30,000/ QALY
CZP **(Deterministic)**	7.024	£93,807	0.279	£3,953	**£14,185**	-	-
TOC(SC) **(Deterministic)**	7.528	£117,033	0.505	£23,226	**£46,018**	-	-
CZP **(Probabilistic)**	7.070	£94,311	0.289	£3,988	**£13,784**	95.36	93.48
TOC **(Probabilistic)**	7.561	£117,142	0.491	£22,832	**£46,501**	0.00	0.16
TOC(IV) **(Probabilistic)**	7.566	£126,323	0.005	£9,181	**£1,945,969**	0.00	0.00

† Rest of the sequence: LEF, GLD, CIC, AZA, PC ‡ CiC PAS not included bio = biosimilar
Source Adapted from ERG report Table 62 and 63

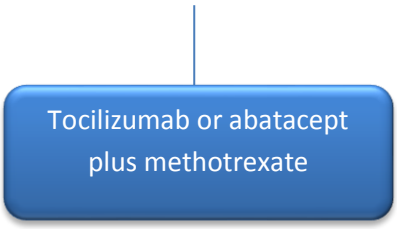
6 Key issues for consideration

Decision Problem

1. Abatacept positioning in clinical practice

The Company has included abatacept plus methotrexate in the third (comparator sequence) and fourth line (with certolizumab) treatment sequences in population A based on their consultation with experts and interpretation of clinical practice in the UK. The ERG consulted its clinical advisors and say the treatment sequences in population A do not reflect clinical practice. It suggested that after rituximab plus methotrexate, tocilizumab plus methotrexate or abatacept plus methotrexate were used, but not both. The ERG note that tocilizumab plus methotrexate after rituximab plus methotrexate is more in line with the recommendation made in TA375 (see page 141 of the ERG report; treatment sequence table shown below following explanation of other changes to the population A sequence).

Figure 8 ERGs interpretation of clinical practice in the UK; represented in the final block from the central pathway from Figure 1 above.



- What is the Committees view on the use of abatacept in clinical practice? Does it align with the Company's sequencing in population A?

2. Best supportive care as a valid comparator.

The Company's interpretation of clinical practice is that BSC is not part of current NICE recommendations for TNF inhibitors and that there is limited

evidence to support such a comparative analysis. The ERGs clinical advisors agreed with this assessment.

- Is best supportive care a valid comparator? And should it have been included in the analysis?

Clinical effectiveness

3. Direct meta-analysis, network meta-analysis and adjusted indirect treatment comparisons for efficacy of combination certolizumab therapy vs monotherapy and comparative efficacy to other bDMARDs

The analysis conducted by the Company can be seen above but the ERG noted several limitations to the Company's approach. It has concerns over the implementation of the fixed and random effects models for the direct meta-analysis and indirect and mixed treatment comparisons. The methods used to estimate treatment effects in the NMA usually require at least 5 studies to prevent an overestimation and inflation of results and a fixed- effects model is deemed inappropriate as it does not allow for heterogeneity to be modelled, which is expected here. It is wrong to ignore heterogeneity on the basis of the treatment effects from different analyses being in the same direction (see ERG report section 4.3, page 96 for full details).

- Do the meta-analyses provide a reliable estimate of treatment effect of certolizumab pegol compared to other bDMARDs?

- What does the data suggest about comparative effectiveness of certolizumab pegol with rituximab?

Cost effectiveness

4. Treatment sequence for population A

In population A the Company defined the intervention and comparator sequences based on the Company's interpretation of previous NICE guidelines and clinical guidelines in England and Wales. The Company's intervention sequence only differs from the comparator sequence by the inclusion of an extra line of therapy; certolizumab plus methotrexate (see Table 6 above). The ERG considered that the Company's approach for population A reflected an elongated sequence rather than a comparison of certolizumab plus methotrexate with rituximab plus methotrexate, as per the NICE scope (see page 140 of the ERG report). However it considered that there were sequences missing from a fully incremental analysis, since certolizumab plus methotrexate could be considered after rituximab plus methotrexate, and certolizumab plus methotrexate could replace rituximab plus methotrexate but neither of these treatment options had been considered. The ERG therefore included these in its preferred base case.

- What is the Committees preferred approach to modelling population A? Does either approach effectively show where certolizumab pegol fits in the main pathway?

5. Inclusion of different biosimilars and formulations in the treatment sequences.

The Company included these in their sensitivity analysis (see above). The ERG believed that the inclusion of these biosimilars and formulation should be part of the base case, not just limited to the sensitivity analysis. It included the sub-cutaneous (SC) formulations of abatacept and tocilizumab, infliximab biosimilars and Benepali (a new etanercept biosimilar) as comparators in its analysis. Benepali is administered weekly as a 50mg/ml solution for injection

in a pre-filled syringe or pre-filled pen. The cost to the NHS of each dose reported in MIMS[86] (in May 2016) is £164.00. For the ERGs analysis it included the SC formation tocilizumab for population A. in population B and C, the ERG used the Company original sequences with the inclusion of biosimilars for infliximab and etanercept and the SC formulations for tocilizumab and abatacept, in population B. In population C it included the biosimilar for etanercept and the SC formulation for tocilizumab. Of the incremental analysis (ICERs shown above in the ERGs preferred base case results) the analysis shows that tocilizumab(SC) ICERs are higher in populations B and C but these should be interpreted with caution as the confidential PAS for tocilizumab have not been incorporated (see ERG confidential comparator PAS appendix).

Table 16 ERG's exploratory analysis amended treatment sequences for Population A

	Sequence name			
	Certolizumab before rituximab	**Certolizumab after rituximab**	**Certolizumab instead of rituximab**	**Rituximab**
Frist	CZP + MTX	RTX + MTX	CZP + MTX	RTX + MTX
Second	RTX + MTX	CZP + MTX	TOC + MTX	TOC + MTX
Third	TOC(SC) + MTX	TOC(SC) + MTX	M + H + S	M + H + S
Fourth	M + H + S	M + H + S	NBT	NBT
Fifth	NBT	NBT	Palliative care	Palliative care
Sixth	Palliative care	Palliative care		

NBT = Non-biologic treatment: a weighted mix of leflunomide, gold, ciclosporin, azathioprine (25% each)

M + H + S = Methotrexate + Hydroxychloroquine + Sulfasalazine. *Source ERG report Table 57*

- What is the Committees view on the inclusion of biosilimars and the sub-cutaneous formulations in the Company's sensitivity analysis? Are they relevant to clinical practice and therefore should they be included in the base case?

6. Network meta-analysis to estimate EULAR response to first treatment and excluding J-RAPID from analysis

The Company used data from the original fixed NMA to estimate these (see above). The ERG would like to have seen a random-effects model for the NMA to estimate EULAR responses for first response to treatment and consider the Company method to be statistically imprecise, not taking into account expected heterogeneity, underestimating uncertainty. The ERG state that using a fixed model in this case is answering the question whether the treatments had an effect in the studies included in the NMA and/or ignoring any potential heterogeneity in treatment effects between studies (see page 142 of the ERG report). The ERG states that there was not a strong enough reasoning for excluding J-RAPID in the Company base case analysis for being a small trial. It states that this should only be excluded if there is evidence to suggest that it is of poor quality.

- Does the NMA provide a reliable estimate for the EULAR response probabilities? What would be the Committees approach?
- Should the J-RAPID trial be included in the network?

7. Comparative effectiveness assumptions for biological DMARDs

In the NMA to estimate EULAR response probabilities the Company had to impose a series of comparative efficacy assumptions, due to lack of comparative data in the target population, in order to carry out the analysis. These were:

- Efficacy of adalimumab, etanercept and infliximab were assumed equivalent to golimumab

- Efficacy of adalimumab and etanercept monotherapy were assumed equivalent to golimumab monotherapy, which was calculated, along with tocilizumab monotherapy, by assuming the relative efficacies of

tocilizumab and golimumab in combination with methotrexate compared to certolizumab plus methotrexate, was maintained when methotrexate was removed. This allowed the efficacy of tocilizumab and golimumab to be calculated from certolizumab monotherapy.

- Infliximab biosimilars were assumed equivalent to infliximab

The ERG considered comparative assumptions in its scenario analysis in which it assumed certolizumab in combination with methotrexate had equal efficacy to infliximab, etanercept and adalimumab, all with methotrexate. It also assumed certolizumab monotherapy in the TNFi-experienced population equal to etanercept and adalimumab monotherapies, going by the reasoning put forward by the Company. This resulted in certolizumab ICER being dominated in population B and C (see section 5.4.2, page 155 of the ERG report for full analysis details). The ERG concluded that there is a lack of data for the efficacy of etanercept, adalimumab and infliximab in combination with methotrexate, and for etanercept and adalimumab monotherapy, in the TNFi-experienced population. The ERG also notes a lack of data for bDMARDs in general for efficacy when failure of two or more TNF inhibitors.

- What is the Committees view on the lack of data for dDMARDs in TNFi-experienced?

- Therefore, what is the Committees' opinion on the placement of certolizumab pegol in the pathway for populations B and C?

8. Retreatment interval for rituximab

Rituximab was considered a drug with irregular administration and Company assumed that re-treatment would occur every 6 months in the base case and explored 9 months in the sensitivity analysis. The ERG prefer to set the rituximab retreatment interval to 7.35 based on the previous appraisal Committee considerations in TA195, which concluded that an interval of 8.7

months was likely an overestimate and an interval of 6 months is not representative of clinical practice. An average retreatment interval was likely to be somewhere in between. The ERG used the midpoint between these two figures: (6+8.7)/2= 7.35. The ERG note that the Company sensitivity analysis of overestimated interval of 9 months noticeable raised the ICER to £49,618/QALY

- What is the appropriate retreatment interval used in clinical practice for rituximab? Is the Company's approach to this reflective of clinical practice?

9. Time to discontinuation on rituximab treatment

The Company's assumed that all bDMARDs have the same time to discontinuation for the first treatment, relying on the approach used in TA195, using the specific parameter values of the Weibull distribution. But the values used in the assessment groups approach in TA195 submission are different for rituximab and abatacept which lead to different treatment duration means (4.06 years for TNFis; 11.31 years for rituximab, and 6.17 years for abatacept). The ERG says that the parameter values should have been that seen in TA195, especially given the importance of its impact.

- What is the Committees view on the Company's assumptions placed on the discontinuation times for bDMARDs? Does it capture the potentially longer duration of treatment with rituximab, given its irregular retreatment intervals?

Key Questions from Considerations

- **Should certolizumab pegol (with or without methotrexate) be recommended as an option at the same point as rituximab? I.e. should it be given as a second anti-TNF after the first one has already failed?**

- **Does the data suggest that certolizumab pegol be not recommended in preference to rituximab which is as effective but less costly?**

7 Authors

Hamish Lunagaria

Technical Lead

Joanne Holden

Technical Adviser

With input from the Evidence review Group and Chair Andrew Stevens

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

Health Technology Appraisal

Certolizumab pegol for treating rheumatoid arthritis after inadequate response to a TNF inhibitor [ID824]

Final scope

Remit/appraisal objective

To appraise the clinical and cost effectiveness of certolizumab pegol within its marketing authorisation for treating rheumatoid arthritis after inadequate response to a TNF inhibitor.

Background

Rheumatoid arthritis is an inflammatory autoimmune disease that typically affects the synovial tissue of the small joints of the hands and feet but can affect any synovial joint, causing swelling, stiffness, pain and progressive joint destruction. It is a systemic disease and can affect the whole body, including the lungs, heart and eyes. Rheumatoid arthritis is usually a chronic relapsing condition which has a pattern of flare-ups followed by periods of lower disease activity; however, for some people, the disease is constantly progressive. Rheumatoid arthritis has a severe impact on quality of life and it is estimated that approximately one-third of people stop work within 2 years because of the disease, and this prevalence increases thereafter.

Estimates of the number of people in England with rheumatoid arthritis vary between about 360,000 and about 690,000. Approximately 15% of people with rheumatoid arthritis have severe disease. Rheumatoid arthritis is about 2 to 3 times more prevalent in women than in men. It can develop at any age, but the usual age of onset in the UK is about 40–70 years.

There is no cure for rheumatoid arthritis and treatment aims to improve quality of life and to prevent or reduce joint damage. Treatment for rheumatoid arthritis usually includes: non-steroidal anti-inflammatory drugs which reduce pain, fever and joint swelling/inflammation, and disease modifying anti-rheumatic drugs (DMARDs). DMARDs may be broadly classed as either non-biological or biological. Non-biological DMARDs include methotrexate, leflunomide and sulfasalazine, while the latter group includes, but is not limited to, tumour necrosis factor (TNF) inhibitors. DMARDs slow the disease process and reduce joint damage. The main aim of management in early disease is to suppress disease activity and induce disease remission, prevent loss of function, control joint damage, maintain pain control and enhance self-management. In established disease, management should address complications, associated comorbidities; and quality of life.

Rituximab in combination with methotrexate is recommended as an option for people with severe rheumatoid arthritis who have had an inadequate response to DMARDs or are intolerant to DMARDs, including a TNF inhibitor (TA195). Abatacept (TA195), adalimumab (TA195), etanercept (TA195), golimumab (TA225), infliximab (TA195) and tocilizumab (TA247) each in combination with methotrexate are recommended as treatment options only if rituximab therapy is contraindicated or is withdrawn because of an adverse event. If the person cannot receive rituximab therapy because they have a contraindication to methotrexate or methotrexate is withdrawn because of an adverse event, adalimumab and etanercept can be given as monotherapy (TA195). If the disease does not respond adequately to 1 or more TNF inhibitors and rituximab, tocilizumab in combination with methotrexate can be given (TA247).

The technology

Certolizumab pegol (Cimzia, UCB Pharma) is an inhibitor of TNF alpha, a pro-inflammatory mediator that is partly responsible for damage to the joints in rheumatoid arthritis. It is administered subcutaneously.

Certolizumab pegol in combination with methotrexate, has a marketing authorisation in the UK for the treatment of moderate to severe, active rheumatoid arthritis in adult patients when the response to DMARDs, including methotrexate, has been inadequate. Certolizumab pegol can be given as monotherapy in case of intolerance to methotrexate or when continued treatment with methotrexate is inappropriate.

Intervention(s)	Certolizumab pegol monotherapy or in combination with methotrexate
Population(s)	Adults with moderate to severe, active rheumatoid arthritis whose disease has not responded adequately to a TNF inhibitor
Comparators	For adults previously treated with other DMARDs including at least 1 TNF inhibitor • Rituximab in combination with methotrexate For adults for whom rituximab is contraindicated or withdrawn • Abatacept, adalimumab, etanercept, golimumab, infliximab and tocilizumab each in combination with methotrexate For adults for whom rituximab therapy cannot be given because methotrexate is contraindicated or withdrawn • Adalimumab monotherapy, etanercept monotherapy or tocilizumab monotherapy

	For people with moderate to severe, active disease despite treatment with biological DMARDs recommended according to NICE guidance • Best supportive care
Outcomes	The outcome measures to be considered include: • disease activity • physical function • joint damage • pain • mortality • fatigue • radiological progression • extra-articular manifestations of the disease • adverse effects of treatment • health-related quality of life.
Economic analysis	The reference case stipulates that the cost effectiveness of treatments should be expressed in terms of incremental cost per quality-adjusted life year. The reference case stipulates that the time horizon for estimating clinical and cost effectiveness should be sufficiently long to reflect any differences in costs or outcomes between the technologies being compared. Costs will be considered from an NHS and Personal Social Services perspective. The availability of any patient access schemes for the intervention or comparator technologies should be taken into account. The availability and cost of biosimilar products should be taken into account.

Other considerations	If evidence allows, the appraisal will consider subgroups of people identified as: • having had primary or secondary failure of response to the first TNF inhibitor; or • having seronegative or seropositive antibody status. If the evidence allows, the appraisal will include the costs of joint replacement therapy and hospital admissions. Guidance will only be issued in accordance with the marketing authorisation. Where the wording of the therapeutic indication does not include specific treatment combinations, guidance will be issued only in the context of the evidence that has underpinned the marketing authorisation granted by the regulator.
Related NICE recommendations and NICE Pathways	Related Technology Appraisals: Technology Appraisal in Preparation, 'Adalimumab, etanercept, infliximab, certolizumab pegol, golimumab, abatacept and tocilizumab for the treatment of rheumatoid arthritis (review of TA guidance 130, 186, 224, 234 and part review of TA guidance 225 and 247)'. Earliest anticipated date of publication TBC. Technology Appraisal No. 247, Feb 2012, 'Tocilizumab for the treatment of rheumatoid arthritis (rapid review of technology appraisal guidance 198)'. Guidance being part reviewed as part of the multiple technology appraisal currently in development. Technology Appraisal No. 280, Apr 2013, 'Abatacept for treating rheumatoid arthritis after the failure of conventional disease-modifying anti-rheumatic drugs (rapid review of technology appraisal guidance 234). Guidance being reviewed as part of the multiple technology appraisal currently in development. Technology Appraisal No. 225, Jun 2011, 'Golimumab for the treatment of rheumatoid arthritis after the failure of previous disease-modifying anti-rheumatic drugs. Guidance being part reviewed as part of the multiple technology appraisal currently in development. Technology Appraisal No. 195, Aug 2010, 'Adalimumab, etanercept, infliximab, rituximab and abatacept for the treatment of rheumatoid arthritis after the failure of a TNF inhibitor'. Transferred to the static list in September 2013. Technology Appraisal No. 186, Feb 2010, 'Certolizumab

	pegol for the treatment of rheumatoid arthritis'. Guidance being reviewed as part of the multiple technology appraisal currently in development. Technology Appraisal No. 130, Oct 2007, 'Adalimumab, etanercept and infliximab for the treatment of rheumatoid arthritis'. Guidance being reviewed as part of the multiple technology appraisal currently in development. Related Guidelines: Clinical Guideline No. 79, Original publication date Feb 2009 (partially updated December 2015) 'Rheumatoid arthritis: The management of rheumatoid arthritis in adults'. Related Quality Standards: Quality Standard No. 33, Jun 2013, 'Quality standard for rheumatoid arthritis'. Review Proposal Date unknown. http://www.nice.org.uk/guidance/qualitystandards/qualitystandards.jsp Related NICE Pathways: NICE Pathway: Rheumatoid arthritis, Pathway created: Jun 2013. http://pathways.nice.org.uk/pathways/rheumatoid-arthritis
Related National Policy	NHS England: NHS England Manual for prescribed specialised services 2013/14. Section 5: Adult highly specialist rheumatology services. NHS England & BMJ Group. Shared Decision Making Sheets: Rheumatoid Arthritis. NHS England. A13. Specialised Rheumatology. National programmes of care and clinical reference groups. National Service Frameworks: Older People Department of Health: Department of Health (2013) NHS Outcomes Framework 2014-2015

Appendix C

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

Single Technology Appraisal (STA)

Certolizumab pegol for treating rheumatoid arthritis after inadequate response to a TNF inhibitor [ID824]

Matrix of consultees and commentators

Consultees	**Commentators (no right to submit or appeal)**
Company • UCB Pharma (certolizumab pegol) Patient/carer groups • Action on Pain • Arthritis Action • Arthritis & Musculoskeletal Alliance (ARMA) • Arthritis Care • BackCare • Disability Rights UK • Equalities National Council • Leonard Cheshire Disability • Muslim Council of Britain • National Rheumatoid Arthritis Society • Pain Concern • Pain Relief Foundation • Pain UK • South Asian Health Foundation • Specialised Healthcare Alliance Professional groups • British Geriatrics Society • British Health Professionals in Rheumatology • British Institute of Musculoskeletal Medicine • British Orthopaedic Association • British Pain Society • British Society for Rheumatology • British Society of Rehabilitation Medicine • Physiotherapy Pain Association • Primary Care Rheumatology Society • Royal College of General Practitioners • Royal College of Nursing	General • Allied Health Professionals Federation • Board of Community Health Councils in Wales • British National Formulary • Care Quality Commission • Department of Health, Social Services and Public Safety for Northern Ireland • Healthcare Improvement Scotland • Medicines and Healthcare products Regulatory Agency • National Association of Primary Care • National Pharmacy Association • NHS Alliance • NHS Commercial Medicines Unit • NHS Confederation • Scottish Medicines Consortium Comparator companies • AbbVie (adalimumab) • Amdipharm Mercury (methotrexate) • Bristol-Myers Squibb (abatacept) • Hameln (methotrexate) • Hospira UK (infliximab biosimilar, methotrexate) • Napp (infliximab biosimilar) • Medac (methotrexate) • Merck Sharp & Dohme (golimumab, infliximab) • Orion(methotrexate) • Pfizer (etanercept, methotrexate) • Roche (rituximab, tocilizumab) • Rosemont (methotrexate) • Sandoz (methotrexate)

Consultees	**Commentators (no right to submit or appeal)**
• Royal College of Pathologists • Royal College of Physicians • Royal Pharmaceutical Society • Royal Society of Medicine • UK Clinical Pharmacy Association <u>Others</u> • Department of Health • NHS England • NHS South Eastern Hampshire CCG • NHS South Kent Coast CCG • Welsh Government	<u>Relevant research groups</u> • Arthritis Research UK • Chronic Pain Policy Coalition • Cochrane Musculoskeletal Group • MRC Clinical Trials Unit • National Institute for Health Research • Society for Back Pain Research <u>Associated Public Health Groups</u> • Public Health England • Public Health Wales

NICE is committed to promoting equality, eliminating unlawful discrimination and fostering good relations between people who share a protected characteristic and those who do not. Please let us know if we have missed any important organisations from the lists in the matrix, and which organisations we should include that have a particular focus on relevant equality issues.

PTO FOR DEFINITIONS OF CONSULTEES AND COMMENTATORS

Definitions:

Consultees

Organisations that accept an invitation to participate in the appraisal; the company that markets the technology; national professional organisations; national patient organisations; the Department of Health and the Welsh Government and relevant NHS organisations in England.

The company that markets the technology is invited to make an evidence submission, respond to consultations, nominate clinical specialists and has the right to appeal against the Final Appraisal Determination (FAD).

All non-company consultees are invited to submit a statement[1], respond to consultations, nominate clinical specialists or patient experts and have the right to appeal against the Final Appraisal Determination (FAD).

Commentators

Organisations that engage in the appraisal process but that are not asked to prepare an evidence submission or statement, are able to respond to consultations and they receive the FAD for information only, without right of appeal. These organisations are: companies that market comparator technologies;
Healthcare Improvement Scotland; other related research groups where appropriate (for example, the Medical Research Council [MRC], National Cancer Research Institute); other groups (for example, the NHS Confederation, NHS Alliance and NHS Commercial Medicines Unit, and the *British National Formulary*.

All non-company commentators are invited to nominate clinical specialists or patient experts.

[1]Non-company consultees are invited to submit statements relevant to the group they are representing.

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

Single technology appraisal

Certolizumab Pegol for Moderate to Severe Rheumatoid Arthritis Inadequately Responding to a Prior TNFi [ID824]

Company evidence submission

Submitted by UCB

Please therefore underline all confidential information, and separately highlight information that is submitted as 'commercial in confidence' in turquoise and all information submitted as 'academic in confidence' in yellow.

February 2016

File name	Version	Contains confidential information	Date
		Yes/no	

Instructions for companies

This is the template for submission of evidence to the National Institute for Health and Care Excellence (NICE) as part of the single technology appraisal (STA) process. Please note that the information requirements for submissions are summarised in this template; full details of the requirements for pharmaceuticals and devices are in the user guide.

This submission must not be longer than 250 pages, excluding appendices and the pages covered by this template.

Companies making evidence submissions to NICE should also refer to the NICE guide to the methods of technology appraisal and the NICE guide to the processes of technology appraisal.

Abbreviations used in this submission

Abbreviation	Definition
ABA	abatacept
ACR	American College of Rheumatology
ACR20	20% improvement in the ACR score
ACR50	50% improvement in the ACR score
ACR70	70% improvement in the ACR score
ADA	adalimumab
AEs	adverse events
AIC	Akaike information Criterion
ANCOVA	analysis of covariance model
AR	adverse reaction
AS	ankylosing spondylitis
axSpA	axial spondyloarthritis
BCP	biochemical profile
bDMARDs	biological DMARDs
BHPR	British Health Professionals in Rheumatology society
BNF	British National Formulary
BSR	British Society for Rheumatology
BSRBR	BSR Biologics Registers
CCP	cyclic citrullinated peptide
CD	Crohn's disease
CDAI	clinical disease activity index
cDMARD	conventional DMARD
CEM	cost effectiveness model
CI	confidence interval
COX	cyclooxygenase
CRD	Centre for Reviews and Dissemination
CRP	C-reactive protein
CV	coefficient of variation
CXR	chest X-ray
CZP	certolizumab pegol
DALYs	disability adjusted life years
DAPS	Direct Access, Pathology services
DAS28	disease activity score 28
D-L	DerSimonian and Laird
DMARDs	disease-modifying antirheumatic drugs
DMARD-IR	DMARD inadequate response
DoF	Data on File

Abbreviation	Definition
DSU	decision support unit
EPAR	European public assessment report
EQ-5D	EuroQoL quality of life scale 5 dimensions
ER	event rate
ERAS	Early RA Study
ESR	erythrocyte sedimentation rate
ETA	etanercept
EULAR	European League Against Rheumatism
Fab'	humanised form of the antigen-binding fragment
FAS	full analysis set
FBC	full blood count
Fc	fragment crystallisable
GBP	Great British Pounds
GI	gastrointestinal
GOL	golimumab
GP	general practitioner
HAQ	health assessment questionnaire
HAQ-DI	HAQ disability index
HRQoL	health-related quality of life
HS	health state
ICER	incremental cost effectiveness ratio
IFX	infliximab
IGWiG	Institute for Quality and Efficiency in Health Care
IPS	Individual Patient Simulation
IR	incidence rate
ISPOR	International Society for Pharmacoeconomics and Outcomes Research
ITC	indirect treatment comparison
ITT	intention-to-treat
LCI	lower confidence interval
LDA	low disease activity
LOCF	last observation carried forward
LYG	life years gained
MACE	major adverse cardiovascular event
mACR	modified ACR
MCID	minimal clinically important difference
MCS	mental component score
MDA	minimal disease activity
MD-HAQ	multidimensional health assessment questionnaire

Abbreviation	Definition
MedDRA	Medical Dictionary for Regulatory Activities
M-H	Mantel-Haenszel
MRI	magnetic resonance imaging
MTA	multiple technology appraisal
MTC	mixed treatment comparison
MTX	methotrexate
n	number of respondents
N	number of patients included in the analysis
N/A	not applicable
NAO	National Audit Office
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
NOAR	Norfolk Arthritis Register
NMR	nuclear magnetic resonance
NRAS	National Rheumatoid Arthritis Society
nr-axSpA	non-radiographic axSpA
NRI	non-responder imputation
ns	not significant
NSAIDs	non-steroidal anti-inflammatory drugs
OI	opportunistic infection
OLE	open-label extension
OR	odds ratio
PAS	patient access scheme
PBO	placebo
PbR	payment by results
PCS	physical component score
PEG	polyethylene glycol
PPD	purified protein derivative
PsA	psoriatic arthritis
PSS	Personal Social Services
pts	patients
PY	patient years
Q2W	every 2 weeks
Q4W	every 4 weeks
QoL	quality of life
QALY	quality-adjusted life year
RA	rheumatoid arthritis
RCTs	randomised controlled trials
RR	risk ratio

Abbreviation	Definition
RTX	rituximab
SAE	serious adverse event
SC	subcutaneous
SD	standard deviation
SDAI	simple disease activity index
SEER	surveillance epidemiology and end results
SF-36	short form 36 questions
SIE	serious infectious event
SJC	swollen joint count
SLR	systematic literature review
SmPC	summary of product characteristics
SPI	sleep problem index
SRQ	Swedish Rheumatology Quality Register
STA	single technology appraisal
TB	tuberculosis
TJC	tender joint count
TNF	tumour necrosis factor
TNFi	TNF inhibitor
TNFi-IR	TNFi inadequate response
TOC	tocilizumab
U	urinalysis
UCI	upper confidence interval
WHO	World Health Organisation
WPS	workplace productivity survey
WPS-RA	workplace productivity survey for RA
WTP	willingness-to-pay

Contents

Tables and figures

Tables

Figures

1 Executive summary

1.1 *Statement of decision problem*

Table 1: The decision problem

	Decision problem addressed in the company submission	Final scope issued by NICE	Rationale if different from the final NICE scope
Population	Adults with moderate to severe active rheumatoid arthritis whose disease has not responded adequately to a tumour necrosis factor (TNF) inhibitor (TNFi). Moderate to severe disease activity is defined as disease activity score 28 (DAS28)>3.2. This population was selected based on British Society for Rheumatology (BSR) recommendations that initiating a first TNFi should be done in patients with moderate to severe disease,[1] and initiating a second line biologic therapy following previous TNFi treatment does not require disease severity as a criterion for eligibility.[2]	Adults with moderate to severe, active rheumatoid arthritis whose disease has not responded adequately to a TNFi	N/A
Intervention	CIMZIA® (certolizumab pegol, CZP) monotherapy or in combination with methotrexate (MTX).	CZP monotherapy or in combination with MTX.	N/A
Comparator (s)	For adults previously treated with other disease-modifying antirheumatic drugs	For adults previously treated with other DMARDs including at least 1 TNFi	The 3rd bulleted patient group indicates that TOC monotherapy should be

	Decision problem addressed in the company submission	Final scope issued by NICE	Rationale if different from the final NICE scope
	(DMARDs) including at least 1 TNFi • Rituximab (RTX) in combination with MTX For adults for whom RTX is contraindicated or withdrawn • Abatacept (ABA), adalimumab (ADA), etanercept (ETA), golimumab (GOL), infliximab (IFX) and TOC each in combination with MTX For adults for whom RTX therapy cannot be given because MTX is contraindicated or withdrawn. • ADA monotherapy, ETA monotherapy or TOC monotherapy	• RTX in combination with MTX For adults for whom RTX is contraindicated or withdrawn • ABA, ADA, ETA, GOL, IFX and TOC each in combination with MTX For adults for whom RTX therapy cannot be given because MTX is contraindicated or withdrawn. • ADA monotherapy, ETA monotherapy or TOC monotherapy For people with moderate to severe, active disease despite treatment with biological DMARDs recommended according to NICE guidance • Best supportive care	considered a comparator. We would like to note that as per current NICE Pathways for Drug Treatment for RA (26th March 2015), and NICE TA247 (22nd February 2015) TOC monotherapy is not recommended as a treatment option in this group of patients. Nevertheless, to address the final scope of this appraisal, TOC has been included in the economic evaluation. The 4th bulleted patient group as per the scope issued by NICE does not reflect the current NICE recommendations for TNFis (NICE Pathways for Drug Treatment for RA (26th March 2015); NICE commissioning algorithm (May 2013)). Furthermore, limited evidence supports the evaluation of CZP within this patient group. As such the submission does not include any assessment of CZP in the 4th population.
Outcomes	Patient Reported Outcomes (PROs) are critically important to understanding the outcomes of a treatment. Studies utilising these measures have shown RA to have a considerable impact on quality of life. Patients suffer from pain, fatigue, limitations to physical function and disability, as well as experiencing effects	The outcome measures to be considered include: • Disease activity • Physical function • Joint damage • Pain • Mortality	N/A

	Decision problem addressed in the company submission	Final scope issued by NICE	Rationale if different from the final NICE scope
	on their psychological, social and emotional well-being. Furthermore, impairment due to the disease on workplace activities and within household have also been reported.	• Fatigue • Radiological progression • Extra-articular manifestations of the disease • Adverse effects of treatment • Health-related quality of life	
Economic analysis	The reference case stipulates that the cost effectiveness of treatments should be expressed in terms of incremental cost per quality-adjusted life year. Due to the chronic nature of RA and, consequently, the lifelong nature of its treatment and associated costs a lifetime horizon will be used. Costs were considered from an NHS and Personal Social Services perspective. Since other patient access schemes (PASs) agreed with the Department of Health for comparators are not in the public domain, analyses using published prices were conducted. Sensitivity analyses on these results were then carried out with a range of discounts in order to capture the confidential PAS for competitors.Unit costs for the biosimilars of IFX were based on NHS published	The reference case stipulates that the cost effectiveness of treatments should be expressed in terms of incremental cost per quality-adjusted life year. The reference case stipulates that the time horizon for estimating clinical and cost effectiveness should be sufficiently long to reflect any differences in costs or outcomes between the technologies being compared. Costs will be considered from an NHS and PSS perspective. The availability of any patient access schemes for the intervention of comparator technologies should be taken into account. The availability and cost of biosimilar products should be taken into account.	As other patient access schemes for comparators are not in the public domain analyses using published prices will be conducted. Sensitivity analyses on these results will then be carried out with a range of discounts in order to capture the confidential patient access schemes/agreements for competitors. Unit costs for the biosimilars of IFX were based on NHS published prices, to ensure consistency in the costing approach for all comparators, as well as consistency with the NICE Guidance to the methods of technology appraisal (2013; Section 5.5.2).

	Decision problem addressed in the company submission	Final scope issued by NICE	Rationale if different from the final NICE scope
	prices.		
Subgroups to be considered	N/A	If evidence allows, the appraisal will consider subgroups of people identified as: • Having had primary or secondary failure of response to the first TNFi; or • Having seronegative or seropositive antibody status. If the evidence allows, the appraisal will include the costs of joint replacement therapy and hospital admissions. Guidance will only be issued in accordance with the marketing authorisation. Where the wording of the therapeutic indication does not include specific treatment combinations, guidance will be issued only in the context of the evidence that has underpinned the marketing authorisation granted by the regulator.	There was insufficient evidence available to support robust analyses within the subgroups specified by NICE.
Special considerations including issues related to equity or equality	N/A	N/A	N/A

N/A: not applicable

1.2 *Description of the technology being appraised*

Table 2: Technology being appraised

UK approved name and brand name	Certolizumab pegol (CZP) CIMZIA®
Marketing authorisation/CE mark status	UK marketing authorisation, 2009
Indications and any restriction(s) as described in the summary of product characteristics	*Rheumatoid arthritis* CZP, in combination with MTX, is indicated for: • the treatment of moderate to severe, active RA in adult patients when the response to DMARDs including MTX, has been inadequate. CZP can be given as monotherapy in case of intolerance to MTX or when continued treatment with MTX is inappropriate • the treatment of severe, active and progressive RA in adults not previously treated with MTX or other DMARDs. CZP has been shown to reduce the rate of progression of joint damage as measured by X-ray and to improve physical function, when given in combination with MTX. *Axial spondyloarthritis (axSpA)* CZP is indicated for the treatment of adult patients with severe active axSpA, comprising: *Ankylosing spondylitis (AS)* Adults with severe active AS who have had an inadequate response to, or are intolerant to nonsteroidal anti-inflammatory drugs (NSAIDs). *axSpA without radiographic evidence of AS (nr-axSpA)* Adults with severe active nr-axSpA but with objective signs of inflammation by elevated C-reactive protein (CRP) and /or magnetic resonance imaging (MRI), who have had an inadequate response to, or are intolerant to NSAIDs. *Psoriatic arthritis (PsA)* CZP, in combination with MTX, is indicated for the treatment of active PsA in adults when the response to previous DMARD therapy has been inadequate. CZP can be given as monotherapy in case of intolerance to MTX or when continued treatment with MTX is inappropriate.

	CZP is contraindicated in patients with a hypersensitivity to the active substance or to any of the excipients, active tuberculosis or other severe infections such as sepsis or opportunistic infections, and in patients with moderate to severe heart failure (NYHA classes III/IV).[3]
Method of administration and dosage	Loading dose The recommended starting dose of CZP for adult patients is 400 mg (given as 2 subcutaneous injections of 200 mg each) at Weeks 0, 2 and 4. For RA, MTX should be continued during treatment with CZP where appropriate. Maintenance dose *Rheumatoid arthritis* After the starting dose, the recommended maintenance dose of CZP for adult patients with RA is 200 mg every 2 weeks. Once clinical response is confirmed, an alternative maintenance dosing of 400 mg every 4 weeks can be considered. MTX should be continued during treatment with CZP where appropriate.

1.3 *Summary of the clinical effectiveness analysis*

Clinical effectiveness in patients previously exposed to a TNFi

To date, the clinical evidence for CZP in RA is supported by 10 placebo (PBO)-controlled, randomised clinical trials (RCTs), all of which have provided consistent results, and have supported the rapid onset of action of CZP as well as the sustained efficacy and safety profile of CZP which is consistent with others in the TNFi class. A systematic literature review identified six RCTs comparing CZP in monotherapy or combination with MTX as a treatment of moderate to severe RA in patients with a previously inadequate response or intolerance to TNFi therapy (excluding CZP). The studies compared CZP with PBO (except for one study, PREDICT, which randomised patients to CZP by assessment criteria to be used during the study) for 24 to 52 weeks. Outcomes considered for analysis of clinical efficacy included, amongst others, clinical response (ACR and EULAR responses), disease activity (DAS28(ESR) [disease activity score 28(erythrocyte sedimentation rate)] remission rates and clinical disease activity index [CDAI] scores), physical function (health assessment questionnaire disability index (HAQ-DI), quality of life (short form 36 questions [SF-36]) and workplace productivity (workplace and household productivity survey for RA [WPS-RA]).

All six RCTs (REALISTIC, DOSEFLEX, PREDICT, SWITCH, J-RAPID and HIKARI) provided consistently favourable results for the efficacy of CZP in patients with moderate to severe RA and previous experience to TNFi treatment.[4-9] CZP provided significant, clinically meaningful benefits to patients across a broad spectrum of

clinical and patient-relevant outcomes. The REALISTIC trial, with 1,063 randomised patients receiving either CZP 200 mg Q2W or PBO, included 400 patients with prior TNFi use, revealed comparable ACR response rates in both the overall study population and the TNFi experienced population. By Week 12, the end of the double-blind phase, both ACR20 and ACR50 response rates were significantly higher in the CZP-treated group than the PBO group within the TNFi experienced population.[4]

Similarly, TNFi experienced patients receiving CZP achieved significant reductions in HAQ-DI score, compared to PBO.[4] In the SWITCH study, which exclusively recruited TNFi experienced patients, significant ACR20 response rates were reached by Week 12 of treatment with CZP in combination with conventional DMARDs (cDMARDs) versus cDMARDs.[7] This observation was mirrored by a comparable mACR20 response rate of 61.7% (see Appendix 8.6.2 for a description of mACR) within the TNFi experienced patients of the PREDICT study, treated with open-label CZP at Week 12.[6] Taken together, these results highlight the effectiveness of CZP treatment in RA patients regardless of prior TNFi experience.[4, 6, 7] In line with these observation, the DOSEFLEX study, which randomised patients for treatment with two dosing regimens of CZP, or PBO, after a 16-week run-in phase of CZP in combination with MTX,[5] demonstrated a comparable effect of CZP treatment on ACR improvement in the TNFi experienced and overall study populations during the double-blind phase. Two studies conducted with RA patients in Japan, J-RAPID and HIKARI,[8, 9] contained a small subset of TNFi experienced patients, and showed similar responses to CZP in both TNFi naïve and TNFi experienced patients.

Meta-analysis

No head-to-head studies of CZP and other biological DMARDs have been published to date. Therefore clinical evidence from CZP trials has been indirectly compared to trials of ADA, ABA, ETA, GOL, IFX, RTX and TOC, using MTX as a common comparator.

Meta-analysis results show that CZP in combination with MTX is associated with a higher clinical response, (as measured by ACR response, EULAR response, and DAS28 remission) compared to MTX alone at 3 months. Results are also statistically significant for ACR20, ACR50, and EULAR (good) response. Furthermore, CZP monotherapy is associated with numerically higher clinical response than PBO at 3 months, albeit without statistical significance. These results need to be interpreted cautiously due to the small population sizes of trials J-RAPID and HIKARI, which, next to the larger REALISTIC study, provided data for the meta-analyses.

The indirect analysis conducted showed that CZP in combination with MTX is at least as effective is its comparators in patients with moderate to severe RA, which have been previously exposed to TNFi. The wide credible intervals noted in most of the analysis results reflect the minimal differences in relative clinical effect between

CZP and the comparators considered. There was no published evidence of comparator biologics in monotherapy to enable an indirect comparison vs CZP.

Safety

An updated pooled analysis of safety data pooled across 10 RA RCTs and their open-label extensions (OLEs), totalling to 4,049 CZP-treated patients, revealed no additional safety signals for CZP treatment, as compared to the licensed SmPC. Accordingly, the adverse event (AE) profile for CZP, including serious infectious events (SIEs) and serious adverse events (SAEs), was found to be in line with that previously reported, with comparable safety of CZP in the TNFi experienced population to the overall safety profile for CZP and confirmed by registries and real-life data.[10-12]

Strengths and limitations of the evidence provided

The data presented in this submission comes for a range of high-quality RCTs, including broad, clinically relevant populations of RA patients, closely resembling the mix of patients seen in daily clinical practice.[4] Recruitment of high proportions of patients with prior TNFi experience (or exclusively these patients, in the case of the SWITCH trial), providing robust pre-defined subgroups, provides a substantial amount of relevant, high quality data in alignment with the decision problem. Despite these strengths, limitations of the evidence presented here include a small sample size in some included studies that reduced the external validity of conclusions drawn from these studies and a lack of published head-to-head clinical comparison of CZP with comparators in the relevant patient population.

1.4 *Summary of the cost effectiveness analysis*

A cohort-based Markov state transition model was developed to assess the cost effectiveness of CZP in TNFi inadequate response (TNFi-IR). The structure of the model is similar to other models in RA, including the model submitted for CZP in DMARD inadequate response (DMARD-IR) in TA375 and TA186. This type of model was chosen because it is well suited to modelling the prognosis of patients with chronic diseases where events (such as treatment switching) can recur, and where the costs and benefits of an intervention accumulate over time. The Markov-type model has been successfully applied to evaluate the cost effectiveness of treatments in RA previously. Additionally, the introduction of tunnel states to the model makes it possible to track the time patients spend in certain states, such as the time on subsequent therapy. These states can be used to incorporate time-varying transition probabilities, and to relax the “memoryless” limitation commonly associated with Markov models.

Current recommendations on the use of CZP in England and Wales are subject to a PAS, in which treatment for patients is provided to the NHS free of charge for the

first 12 weeks. The CZP PAS has been accounted for in the cost effectiveness analysis.

Key assumptions within the model are as follows:

- The following assumptions were made during the analyses of initial response to therapy:
    - **All populations:** The comparative efficacy of CZP versus PBO was estimated by comparing (i) the reported EULAR response probabilities for CZP at six months to (ii) estimated response probabilities for PBO derived by mapping response at three months to response at six months
    - **Populations B/C:** Due to the lack of published trial data, ADA, ETA, and IFX were modelled as a treatment class assuming equivalent efficacy to GOL, when given in combination with MTX
    - **Population B:** Biosimiliars to IFX were assumed to have equivalent efficacy to IFX
    - **Population C:** The efficacy of TOC monotherapy was modelled assuming that the comparative efficacy of TOC versus PBO and other therapies, is consistent across both combination and monotherapy regimens
    - **Population C**: Due to the lack of published trial data, the efficacy of comparator TNFis was modelled as a treatment class. The efficacy of ADA and ETA monotherapy were modelled using the effect size estimates for GOL versus CZP, derived from the NMA of biological DMARDs (bDMARDs) given in combination with MTX
- In the base case, the costs of ABA and TOC were calculated assuming administration as an intravenous formulation, on the basis that the clinical efficacy of these therapies was derived from studies where therapy was given via infusion. The cost impact of assuming administration via the subcutaneous formulations of ABA and TOC were considered in the scenario analyses (see Table 101)
- Re-treatment with RTX was assumed to occur every 6 months in the base case analyses and every 9 months in sensitivity analyses

For a complete list of model assumptions, please see Section 5.6.2.

The incremental cost effectiveness results (probabilistic results) are summarised for each population in Table 3, Table 4, and Table 5.

Population A. In the base case analysis, CZP + MTX was associated with an increase in quality-adjusted life years (QALYs) (+0. 291) and an increase in costs (+£9,782) when compared to RTX + MTX, leading to an incremental cost

effectiveness ratio (ICER) of £33,665. The probabilities that CZP was cost effective at thresholds of £20,000 and £30,000 were 2.98% and 37.40%, respectively.

Table 3: Base case incremental cost effectiveness results (probabilistic analysis - Population A (adults previously treated with other DMARDs including at least one TNFi)

Therapy	Mean costs	Difference in costs* (CZP vs comparator)	Mean QALYs	Difference in QALYs* (CZP vs comparator)	ICER (CZP vs comparator)	Probability of cost effectiveness at WTP threshold of £20,000/QALY (%)	Probability of cost effectiveness at WTP threshold of £30,000/QALY (%)
CZP + MTX	£150,413		7.295			2.98%	37.40%
RTX + MTX	£140,631	£9,782	7.005	0.291	£33,665	97.02%	62.60%

Note: * difference based on costs or QALYs for CZP versus comparator.
Results represent mean estimates per patient.
WTP: willingness-to-pay.

Population B. Base case results indicated that CZP + MTX was dominant when compared to ABA + MTX or IFX + MTX. CZP + MTX was cost effective compared to GOL + MTX, ADA + MTX, and ETA + MTX, resulting in ICERs of £3,527, £3,635, and £1,938, respectively. When compared with TOC + MTX, CZP + MTX was less costly but less effective, resulting in an ICER of £129,321 (TOC vs CZP). In the full incremental analysis, CZP + MTX was considered the optimal treatment strategy in Population B at willingness-to-pay (WTP) thresholds between £4k and £130k. The probabilities that CZP + MTX is cost effective at thresholds of £20k and £30k per QALY gained was 99.5% and 95.9%, respectively.

Table 4: Base case incremental cost effectiveness results – probabilistic analysis - Population B (adults for whom RTX is contraindicated or withdrawn)

Therapy	Mean costs	Difference in costs* (CZP vs comparator)	Mean QALYs	Difference in QALYs* (CZP vs comparator)	ICER (CZP vs comparator)	Probability of cost effectiveness at WTP threshold of £20,000/QALY (%)	Probability of cost effectiveness at WTP threshold of £30,000/QALY (%)
CZP + MTX	£98,916		6.302			99.5%	95.9%
ABA + MTX	£116,217	-£17,301	6.087	0.214	CZP dominates	0.0%	0.0%
ADA + MTX	£97,944	£972	6.034	0.267	£3,635	0.2%	1.2%
ETA + MTX**	£98,402	£513	6.037	0.265	£1,938	0.1%	1.1%
GOL + MTX	£97,984	£931	6.038	0.264	£3,527	0.2%	1.7%
IFX + MTX**	£102,272	-£3,356	6.038	0.263	CZP dominates	0.0%	0.0%
TOC + MTX	£125,518	-£26,603	6.507	-0.206	£129,321 (TOC vs CZP)	0.0%	0.0%

Note: * difference based on costs or QALYs for CZP versus comparator
**Original brands only (ie Remicade (IFX) and Enbrel (ETA))
CZP dominates comparator (ie. CZP is more effective and less costly)
Results represent mean estimates per patient.

Population C. The base case analysis indicated that CZP was cost effective compared to ADA (ICER=£4,943) and ETA (ICER=£3,514), and less effective and less costly than TOC (ICER of £129,177 for TOC vs CZP). In the full incremental analysis, CZP was the optimal treatment strategy at thresholds between £5k and £130k. The probabilities that CZP is cost effective at thresholds of £20k and £30k per QALY gained were 99.6% and 97.2%, respectively.

Table 5: Base case incremental cost effectiveness results – probabilistic analysis - Population C (adults for whom RTX therapy cannot be given because MTX is contraindicated or withdrawn)

Therapy	Mean costs	Difference in costs* (CZP vs comparator)	Mean QALYs	Difference in QALYs* (CZP vs comparator)	ICER (CZP vs comparator)	Probability of cost effectiveness at WTP threshold of £20,000/QALY (%)	Probability of cost effectiveness at WTP threshold of £30,000/QALY (%)
CZP	£97,550		6.141			99.64%	97.20%
ADA	£96,198	£1,352	5.868	0.274	£4,943	0.240%	1.700%
ETA**	£96,587	£963	5.867	0.274	£3,514	0.120%	1.100%
TOC	£123,749	-£26,199	6.344	-0.203	£129,177 (TOC vs CZP)	0.000%	0.000%

Note: * difference based on costs or QALYs for CZP versus comparator
**Original brands only (ie Enbrel (ETA))

Conclusions:

In conclusion, the results of the cost effectiveness analysis strongly support the case that CZP is a cost effective treatment option in TNFi-IR:

- for whom RTX is contraindicated or withdrawn
- for whom RTX therapy cannot be given because MTX is contraindicated or withdrawn

In these populations, CZP either dominates existing therapies or is cost effective within the bounds of conventional thresholds (ie. <£30,000).

In adults previously treated with other DMARDs, including at least one TNFi, CZP may be considered a cost effective treatment option, if provided prior to treatment with RTX and standard therapies. Although the base case ICER (£33k) for this population is just above the conventional threshold of £30,000, it should be noted that when taking into account the wider societal impact of RA, the potential for longer duration of CZP therapy compared to RTX therapy, or the modelling of EQ-5D based solely on HAQ, the ICER reduces to below the £30,000 threshold (Section 5.8.4).

No analyses were presented for patients with moderate to severe, active disease despite treatment with bDMARDs.

2 The technology

CZP is a TNFi for the treatment of RA.

- CZP has a novel molecular structure to target TNF-α and is the only polyethylene glycol (PEG)-conjugated (PEGylated) TNFi approved for the treatment of RA.[3]
- CZP has been approved for use in RA in 61 countries worldwide.[13]
- CZP is currently recommended by NICE for use in severe RA for patients who have shown an inadequate response to MTX and another cDMARD.[14]

Benefits of CZP treatment.

- Treatment with CZP is associated with a rapid response in a broad population of patients, naïve or previously exposed to TNFi, allowing for early indication of therapeutic benefit.[4, 6,15]
- Response to CZP within the first 12 weeks of treatment is achieved within a majority of RA patients and highly predictive of clinical outcome at 1 year.[15]
- In addition to clinical benefits, CZP has also been shown to be associated with large benefits in workplace and household productivity, as well as social participation, which can result in economic benefits.[16,17]
- After a loading dose of 400 mg at Weeks 0, 2 and 4, CZP can be administered at the maintenance dose of 200 mg every 2 weeks (Q2W) or alternatively as 400 mg every 4 weeks (Q4W), once clinical response with CZP 200 mg Q2W has been confirmed.

The CZP patient access scheme (PAS) provides 12 weeks of therapy with CZP to the NHS for free.

- The established link between clinical benefits at Week 12 and long-term outcomes of CZP-based therapy allows for an early decision on effectiveness of therapy and avoidance of unnecessary costs.
- The PAS, agreed with the Department of Health, reduces the costs of treatment from £10,367.50 to £6,793 in the first year, with subsequent costs of £9,295 a year thereafter.

2.1 *Description of the technology*

Certolizumab pegol (CIMZIA®, UCB) is a PEGylated, Fc (fragment crystallisable)-free, TNFi for the treatment of RA. It has a high affinity, neutralising potency, and specificity for human TNF-α, a key pro-inflammatory cytokine with a central role in inflammatory processes. CZP is the only PEGylated TNFi approved for the treatment of RA.[3]

2.2 *Marketing authorisation/CE marking and health technology assessment*

CZP, in combination with MTX, is indicated in Europe, for the treatment of moderate to severe active RA in adults when the response to DMARDs, including MTX, has been inadequate. CZP can be given as monotherapy in case of intolerance to MTX or when continued treatment with MTX is inappropriate.[3] CZP received this approval in the UK (as part of the European Union and associated countries [Iceland, Norway and Liechtenstein]) on 1st October 2009.[18] CZP is also indicated for the treatment of

severe, active and progressive RA in adults not previously treated with MTX or other DMARDs.

CZP is also approved within the European Union for the treatment of severe, active axSpA and active PsA. See Appendix 8.1 for full details of current regulatory approval.

CZP is contraindicated in people with active tuberculosis or other severe infections, and in people with moderate or severe heart failure. The summary of product characteristics (SmPC) lists no AEs as very common but notes that in clinical trials the most common AEs were bacterial and viral infections.[3] For full details of AEs and contraindications, please refer to the SmPC.[3]

There were three main issues discussed in the European public assessment report (EPAR) for the approval of CZP for the treatment of RA. At the time of consideration, there was insufficient evidence to support a lack of reproductive and developmental toxicity with CZP. Therefore it was not recommended for use in women of child-bearing age and pregnancy; however this has been adequately reflected in the product information. The issue of opportunistic and serious infections was also raised as these were the most common AEs. The potential risk of malignancy with long-term (longer than 52 weeks) use of CZP was also highlighted, however adequate evidence in the SmPC and an acceptable risk management plan led to the recommendation of CZP for marketing authorisation.[18] For full details of the main issues identified, please refer to the EPAR.[18]

CZP has been approved for use in RA in 61 countries worldwide encompassing Canada and North America, Europe, Latin America, the Middle East and Africa.[13] For full registration details, see Appendix 8.1.

CZP is currently recommended by NICE for use in severe RA for patients who have shown an inadequate response to MTX and another cDMARD for at least 6 months of treatment, or for those patients who experience AEs within the first 6 months of treatment with a different TNFi (TA186).[19] CZP is approved for administration in combination with MTX and also as a monotherapy for patients with a contraindication to MTX.[19] In a recent multiple technology appraisal (MTA; TA375), CZP has been recommended by NICE for the treatment of patients with severe RA, either in combination with MTX, in those whose disease has not responded to intensive therapy with a combination of cDMARDs, or as a monotherapy, for those who cannot take MTX because it is contraindicated or because of intolerance.[14] Furthermore, CZP has been recommended for the treatment of severe active AS and nr-axSpA after unsuccessful therapy with, or contraindicated for, NSAIDs (TA383),[20] and is currently under evaluation as part of an MTA for PsA after treatment with DMARDs (ID579).[21]

In this submission, CZP will be evaluated within its licensed RA indication.[3] The patient population considered in this STA comprises patients with an inadequate response to TNFi therapy.

2.3 *Administration and costs of the technology*

As indicated in the European SmPC, the recommended starting dose of CZP for adult patients, known as the loading dose, is 400 mg (given as 2 subcutaneous (SC) injections of 200 mg each) at Weeks 0, 2 and 4. For RA, MTX should be continued during treatment with CZP where appropriate. After the loading dose, the recommended maintenance dose of CZP for adult patients with RA is 200 mg Q2W. Once clinical response is confirmed, an alternative maintenance dosing of 400 mg Q4W can be considered. MTX should be continued during treatment with CZP where appropriate.[3]

The company has agreed a PAS with the Department of Health. In the scheme, the first 12 weeks of therapy (currently 10 pre-loaded syringes of 200 mg each) with CZP are provided to the NHS free of charge.

The net price of CZP is £357.50 per 200 mg prefilled syringe (British National Formulary [BNF], October 2015) (Table 6). Assuming 3 initial doses of 400 mg at Weeks 0, 2 and 4, followed by maintenance doses of 200 mg Q2W, the cost in the first year with the PAS is £6,793 (or £10,367.50 without the PAS) and then £9,295 per year. Costs may vary in different settings because of negotiated procurement discounts.

The Department of Health considered that the PAS for CZP does not constitute an excessive administrative burden on the NHS.

Table 6: Costs of the technology being appraised

	Cost	Source
Pharmaceutical formulation	CZP is formulated as a solution for SC injection. Each pre-filled syringe (type I glass) with plunger stopper (bromobutyl rubber) contains 200 mg CZP in 1 mL. None of the components of the syringe contain latex. CZP is provided in a pack size of 2 pre-filled syringes and 2 alcohol wipes.	SmPC[3]
Acquisition cost (excluding VAT)	The list price per pack is £715, excluding VAT, which is equivalent to a net price of £357.50 per 200 mg pre-filled syringe. Under the PAS agreed with the Department of Health, the first 12 weeks of CZP are provided free of charge which is equivalent to 10 vials at a total cost saved of £3,575 in Year 1 of treatment.	BNF, October 2015[22]
Method of administration	Subcutaneous injection. Suitable sites for injection would include the thigh or abdomen.	SmPC[3]
Doses	Loading dose of 400 mg at Weeks 0, 2 and 4 then maintenance dose of 200 mg Q2W.	SmPC[3]

Dosing frequency	Loading dose of 400 mg at Weeks 0, 2 and 4 then maintenance dose of 200 mg Q2W, or 400 mg Q4W once clinical response has been confirmed.	SmPC[3]
Average length of a course of treatment	Approximately 4 years for the first agent used in the TNFi experienced population, approximately 20 years on therapy in total.	Generated from the cost-effectiveness model (Section 5)
Average cost of a course of treatment	The cost in the first year with the PAS is £6,793 and then £9,295 per year.	TA375[14]
Dose adjustments	N/A	
Anticipated care setting	Home care	SmPC[3]

2.4 *Changes in service provision and management*

No additional investigations or administrations are required for the technology being appraised. Diagnosis of RA does not depend on the presence of specific biomarkers, therefore no diagnostic tests will be evaluated in this submission.

CZP is administered via SC injection. All CZP patients require one hour of training by a nurse in the administration of SC injections, these patients can then self-administer CZP in the home care setting. Therefore no additional resource use for CZP over other biologics is required.[14]

The resource unit costs of CZP are displayed in Table 7.

Table 7: Resource unit costs for CZP

Item	Unit cost (2015 £)	Source
Rheumatologist visit	£137.00	NHS Reference Costs 14/15: WF01A
GP visit	£65.00	PSSRU 2015 (p. 177, 10.8b)
Nurse visit	£75.00	PSSRU 2015 (p. 172, 10.4)
Hospital day - Pal.	£371.00	PSSRU 2015 (p. 107, 7.1)
IV administration	£173.60	NICE TA247
Full blood count	£3.01	NHS Reference Costs 14/15: DAPS05
Urea and electrolytes	£1.19	NHS Reference Costs 14/15: DAPS04
Liver function test	£3.01	NHS Reference Costs 14/15: DAPS05
Creatinine	£3.01	NHS Reference Costs 14/15: DAPS05
Chest X-ray	£30.23	NHS Reference Costs 14/15: DAPF

GP: general practitioner, Pal: palliative, IV: inteavenous.

Additionally, no additional infrastructure in the NHS is required for the technology being appraised in this submission.

The technology in this appraisal will not affect current clinical monitoring procedures as CZP is already approved for use by NICE, whereby CZP is indicated, either in

combination with MTX or as a monotherapy, where MTX is contraindicated, for patients with severe, active RA who have shown an inadequate response to MTX and another cDMARD (TA375).[14] This present submission proposes an extension to the terms of use of CZP, to include treatment of patients with an inadequate response to a previous TNFi. The definition of no response remains defined as lack of maintenance of a EULAR moderate response after 6 months of treatment.[14]

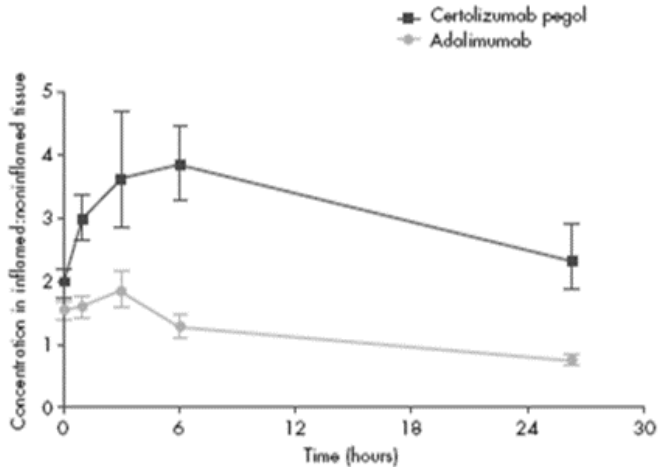
2.5 *Innovation*

2.5.1 CZP structure

CZP has a novel molecular structure. It consists of a humanised form of the antigen-binding fragment (Fab') of an anti-TNF-α monoclonal antibody. The Fc region of this antibody is absent which reduces Fc-mediated responses including antibody dependent cell-mediated cytotoxicity and complement dependent cytotoxicity. In addition, the Fab' fragment has been PEGylated via attachment of a 40 kDa PEG moiety.[23] The addition of PEG increases the stability and half-life of the drug, allowing the molecule to have a half-life comparable to a full antibody (14.4 days).[24] CZP is the only Fc-free, PEGylated Fab' fragment TNFi currently available for the treatment of RA.

PEGylation may also aid retention in inflamed tissue, and CZP has been shown to preferentially accumulate in inflamed tissue in animal models. A study in naïve mice and mice with ongoing collagen-induced arthritis, used fluorescently labelled CZP and ADA to show that both TNFi therapies achieved greater penetration into arthritic paws compared to normal tissue. Importantly, CZP also demonstrated longer and more extensive penetration into inflamed tissue than ADA (Figure 1).[25]

Figure 1: Ratio of ADA and CZP in inflamed and non-inflamed tissue in a mouse model for collagen induced arthritis[25]



Similarly, in humans, radiolabelled CZP has been found to accumulate within inflamed joints within minutes of administration. Furthermore, accumulation of radiolabelled CZP had a high concordance with ultrasound and swollen joint count in patients with RA and peripheral spondyloarthritis.[26]

Although CZP has not been approved for use in pregnancy,[3] pre-clinical and clinical data suggest a lack of active placental transfer of CZP which may be the consequence of the absence of an Fc region. Pre-clinical data showed that anti-TNF-α PEGylated Fab' fragment was undetectable or only detectable at a very low level in rodent foetal samples.[27] An *ex vivo* placental transfer study found that in 6 human placentae, CZP levels in the foetal circulation were consistently below levels of the anti-D IgG control.[28] An independent investigator-driven study measuring placental transfer in 10 pregnant women treated with CZP reported that CZP levels in the cord blood and infant blood on the day of birth were consistently lower than those in maternal blood, suggesting low placental transfer. In some cases, CZP levels in cord and infant blood were below levels of detection.[29, 30] In a recent retrospective analysis of the UCB safety database, analysis of pregnancy outcomes after exposure to CZP supports findings from previous reports, compared to background incidence in the non-exposed population, suggesting a minimal harmful effect of CZP exposure on pregnancy outcomes.[31] The European label for CZP states that there are no adequate data from the use of CZP in pregnant women.[3]

2.5.2 Administration benefits of CZP

In addition to its novel molecular structure, CZP offers a flexible dosing schedule. It can be administered SC, either as a 200 mg Q2W maintenance dose after the initial loading doses have been administered, or alternatively as 400 mg Q4W once a clinical response has been confirmed, depending on patient choice. The 400 mg Q4W option offers more convenient dosing for patients, especially those who frequently travel away from home.

CZP can be self-administered by RA patients. It is supplied as an award-winning, ergonomically-designed pre-filled syringe that was designed in collaboration with OXO Good Grips® to ensure that it is easy to use.[32, 33] This allows convenient self-administration for those patients who feel able to do so.

CZP also offers predictable annual costs. Patients receive the same cumulative dose of CZP regardless of the maintenance dosing regimen and irrespective of their weight: 400 mg over the course of 4 weeks, administered either 200 mg Q2W or 400 mg Q4W.

2.5.3 CZP rapid response

CZP has demonstrated a rapid response in a broad population, including patients who are naïve to TNFi or had been previously exposed to TNFi therapy. In both

REALISTIC and PREDICT, studies including both patients with and without prior TNFi exposure, CZP resulted in a ≥20% improvement in the ACR score (ACR20) as early as Week 2 in approximately one third of patients.[4, 6] This rapid response allows therapeutic benefits to be improved through early decision making and avoidance of unnecessary costs and safety concerns due to avoidable exposure.[15]

2.5.4 The CZP patient access scheme

A primary concern with TNFi treatments is the cost of failure in patients who do not respond to treatment. Based upon the established link between the attained degree of clinical benefit at Week 12 and the long-term outcome of CZP-based therapy, UCB has developed an innovative PAS, in which treatment for patients is provided to the NHS free of charge for the first 12 weeks to allow a clear and early decision on the effectiveness of CZP therapy in the interest of both the NHS and the patient.

2.5.5 Health-related benefits not captured by utility assessments

In addition to clinical benefits, CZP has also been shown to be associated with large benefits in productivity outcomes compared to PBO in a broad population, including patients naïve or previously exposed to TNFi.[16] Productivity outcomes, and the associated economic benefits following biologic treatment, and the positive impact on the patient's state of mind from becoming more productive, are not currently included in the cost per QALY calculation. They are, however, an extremely important aspect of RA treatment in light of the enormous indirect costs associated with RA, and should therefore be considered.[17, 34]

3 Health condition and position of the technology in the treatment pathway

RA is an incurable disease with high patient health burden.

- Approximately 0.8% of the UK population are affected by RA, with 12,000 new cases diagnosed each year.[35, 36]
- Life expectancy of RA patients is decreased by 10 years, with infection, heart disease and gastrointestinal bleeding responsible for the majority of excess mortality.[37]

Poor patient health due to RA has a large impact on the health system and economy.

- In addition to the considerable burden on patients' quality of life, RA also has a substantial economic impact with high indirect costs that can outweigh direct treatment costs.[37-39]
- Typically, the economic burden of RA increases with greater disease activity, as this has a larger disabling effect which impedes productivity.
- In a recent National Rheumatoid Arthritis Society (NRAS) survey 50.1% of surveyed RA patients were unemployed, of these patients, 30.4% had retired due to their RA and 25.1% were currently unable to work due to their RA.[40]

RA patients refractory to their first TNFi therapy have various second line treatment options in the UK.

- An estimated 2–3.5% of RA patients in England have severe disease refractory to conventional treatment who may be eligible for TNFi therapy.[41]
- Not all patients respond adequately to their first TNFi, with 12% of patients in the UK stopping their first TNFi after 6 months due to a lack of efficacy, according to a large (6,379 patient) analysis.[42]
- An inadequate response to a prior TNFi does not predict an inadequate response to a subsequent TNFi in RA patients, as for 74% of patients a second TNFi was effective for at least 6 months.[42]
- All licensed biologic therapies approved for use in RA in the UK in combination with MTX and as monotherapy (except CZP) are recommended for use in inadequate responders to TNFi therapy.[43-45]

Despite having shown similar efficacy to other approved TNFi agents, CZP has not yet been appraised for use in adults with moderate to severe, active RA whose disease has not responded adequately to TNFi therapy in the UK.

- EULAR guidelines for the treatment of RA state that no TNFi is better than any other in patients with active disease, despite initial treatment with a TNFi.[46]
- In the current NICE clinical pathway, CZP is only available as a first-line TNFi and an approach to match the availability of CZP with the rest of the TNFi class is deemed appropriate.[14]

3.1 *Disease Overview*

RA is a systemic chronic inflammatory autoimmune disease that typically affects synovial joints (such as those in the hands and feet), causing swelling, stiffness, pain and progressive, irreversible joint destruction in the absence of adequate treatment.[47] Disease can also occur outside the joints, affecting other organs, including the lungs, heart and eyes.[48] Commonly, the disease course is chronic, relapsing and with a pattern of flare-ups followed by periods of low disease activity.

RA is associated with increased mortality and increasing disability, which has a severe effect on quality of life. It is associated with substantial costs: direct costs of drug acquisition and hospitalisation and indirect costs of reduced productivity.[49, 50]

There are estimated to be approximately 580,000 people with RA in England.[49] Of these, approximately 2–3.5% have severe disease that has not responded to

conventional treatment, who may be eligible to TNFi therapy.[41] RA is 2–3 times more prevalent in women than in men. It can develop at any age, but the peak age of onset in the UK is 40–60 years.[47]

There is no cure for RA. In early disease, management aims to suppress disease activity and induce remission, prevent loss of function, control joint damage, control pain and enhance self-management. In established disease, management should address complications and associated comorbidity, as well as the effect of the condition on the person's quality of life.[51]

Treatment for RA usually includes NSAIDs or cyclooxygenase (COX) 2 inhibitors, and DMARDs. NSAIDs and COX2 inhibitors reduce pain, fever, joint swelling and inflammation (glucocorticoids may also be used to control inflammation), whilst DMARDs slow the disease process and reduce joint damage. DMARDs can include drugs such as MTX, leflunomide and sulfasalazine (referred to as cDMARDs). Also available are a group of drugs including monoclonal antibodies and soluble receptors that modify the disease process by blocking key protein messenger molecules (such as cytokines eg. TNF-α) or cells (such as B lymphocytes).[52] Such drugs are referred to as bDMARDs. In some patients the disease may not respond to either bDMARDs or cDMARDs and for others the response to DMARDs often reduces over time.[42, 53, 54] Therefore, a sequence of treatments is required.

3.2 *The burden of disease*

Approximately 0.8% of the UK population are affected by RA, with 12,000 new cases diagnosed each year.[35, 36] Approximately one third of patients discontinue work within 2 years and at least 10% of RA patients become severely disabled, despite full treatment.[55]

Life expectancy of RA patients is decreased by 10 years, with infection, heart disease and gastrointestinal bleeding responsible for the majority of excess mortality. This excess mortality is attributable to drug treatments for RA as well as the underlying disease and malignancy.[37]

Morbidity associated with RA represents approximately 0.8% of all Disability Adjusted Life Years (DALYs) lost in Europe, with pain, fatigue, functional impairment and depression having a substantial impact on quality of life in RA patients.[50, 56]

In addition to the considerable burden on patients' quality of life, RA also has a substantial economic impact.[37] While there is variation in RA cost estimates between sources, the economic burden of RA is driven largely by indirect costs that can outweigh direct costs.[37-39] These indirect costs are primarily a result of the disability associated with RA that results in both paid and unpaid productivity losses, and can have a profound impact on the lives of patients and their families.[57] Presenteeism (reduced productivity while at work) is a key contributor to the indirect cost of RA,

although absenteeism (absence from work) and reductions in household productivity are also significant.[58] In a recent NRAS survey, 50.1% of surveyed RA patients were unemployed. Of these patients, 30.4% had retired due to their RA and 25.1% were currently unable to work due to their RA. This highlights the impact of RA on patients' productivity and is supported by the finding that 38.0% and 33.7% of respondents reported that their RA was not well controlled enough for them to continue to carry out valued day-to-day activities, or was affecting their confidence and independence, respectively.[40] In a 2007 NRAS survey of RA and employment, it was found that of the people that had given up employment or retired early as a result of their RA, 28.4% did so within one year of diagnosis, and 59% did so within six years,[59] highlighting the rapid loss of long-term productivity from these patients.

In 2009, a National Audit Office (NAO) and Public Accounts Committee report, on the efficiency and effectiveness of services for people with RA in England, estimated the annual healthcare costs to the NHS due to RA to be £560 million. Further, the additional costs to the economy due to sick-leave and work-related disability were estimated at £1.8 billion.[34] An NRAS report published in March 2010 estimated that the overall cost of productivity losses due to RA to the UK economy is almost £8 billion per year.[49] The total productivity losses associated with existing RA cases totalled nearly £52 billion.[49] Typically, the economic burden of RA increases with greater disease activity, as this has a larger disabling effect which impedes productivity.

3.2.1 Patients having received prior TNF inhibitor therapy

Although treatment with TNFi therapies can reduce the burden of disease on both patients and carers, not all patients respond adequately to their first TNFi. The proportion of patients who fail to achieve an adequate response to treatment; defined as a DAS28 score of <2.6 after 6 months on an initial first-generation TNFi (ETA, ADA or IFX) has been shown to include 12–52% of patients treated with TNFi agents and so represents a considerable proportion of the RA population.[42, 60] Within this patient group, a physician survey indicated over 94% of physicians prescribed an alternative TNFi after their patient experienced an inadequate response or AEs with an initial TNFi.[61]

In a large prospective cohort analysis of 6,739 UK active RA patients, 12% stopped their first TNFi after 6 months due to a lack of efficacy.[42] Of these patients, 60% went on to receive a second TNFi. The second TNFi was effective for at least 6 months in 74% of patients, however, at this second-line, a comparable proportion of patients (13%) again discontinued due to lack of efficacy.[42] This suggests that an increased number of treatment options available for patients who have previously failed to respond to at least one TNFi may be beneficial, especially in cases where inadequate response has been demonstrated for more than one prior TNFi.

This is supported by a number of RCTs, which have suggested that an inadequate response to a prior TNFi does not predict an inadequate response to a subsequent TNFi in RA patients.[4-7, 62] In these studies a total of 333 RA patients with secondary failure, defined as a loss of response after initial improvement, on at least one TNFi, were randomised to receive an alternative TNFi (CZP [n=320] or IFX [n=13]). Overall, the majority of patients responded to the new TNFi and in the case of CZP, the response rate in these patients (47.2% ACR20 responders) was similar to the response rate in patients who had not received a prior TNFi (53.5% ACR20 responders). Additionally, the CZP response rates were significantly higher than patients who had been randomised to PBO (27.5% ACR20 responders on placebo after prior TNFi) ($p<0.01$).[4] In patients who received IFX following an inadequate response to ETA, patients receiving second-line IFX showed improved response rates (61.5% ACR20 responders) in comparison to patients who remained on ETA (28.6% ACR20 responders), which highlights the value of alternative TNFi therapy in this group of patients.[62]

Combined, these results highlight a substantial proportion of the RA population who may experience an initial response to a TNFi but subsequently lose this response. These patients may then suffer an increase in burden of disease that may not be improved until initiation of an alternative TNFi therapy, which has a substantial probability of providing therapeutic benefit.

3.3 *RA treatment pathway*

Currently, according to the NICE pathway, moderate to severely active RA is first treated with cDMARDs, including MTX, as shown in the treatment algorithm in Figure 2. If a patient does not exhibit an adequate response to these, bDMARDs are recommended. In the case where a patient may experience no AEs, but either loses response to (secondary failure) or fails to initially adequately respond to their first bDMARD (primary failure), the CD20-inhibitor RTX, in combination with MTX, is currently recommended.

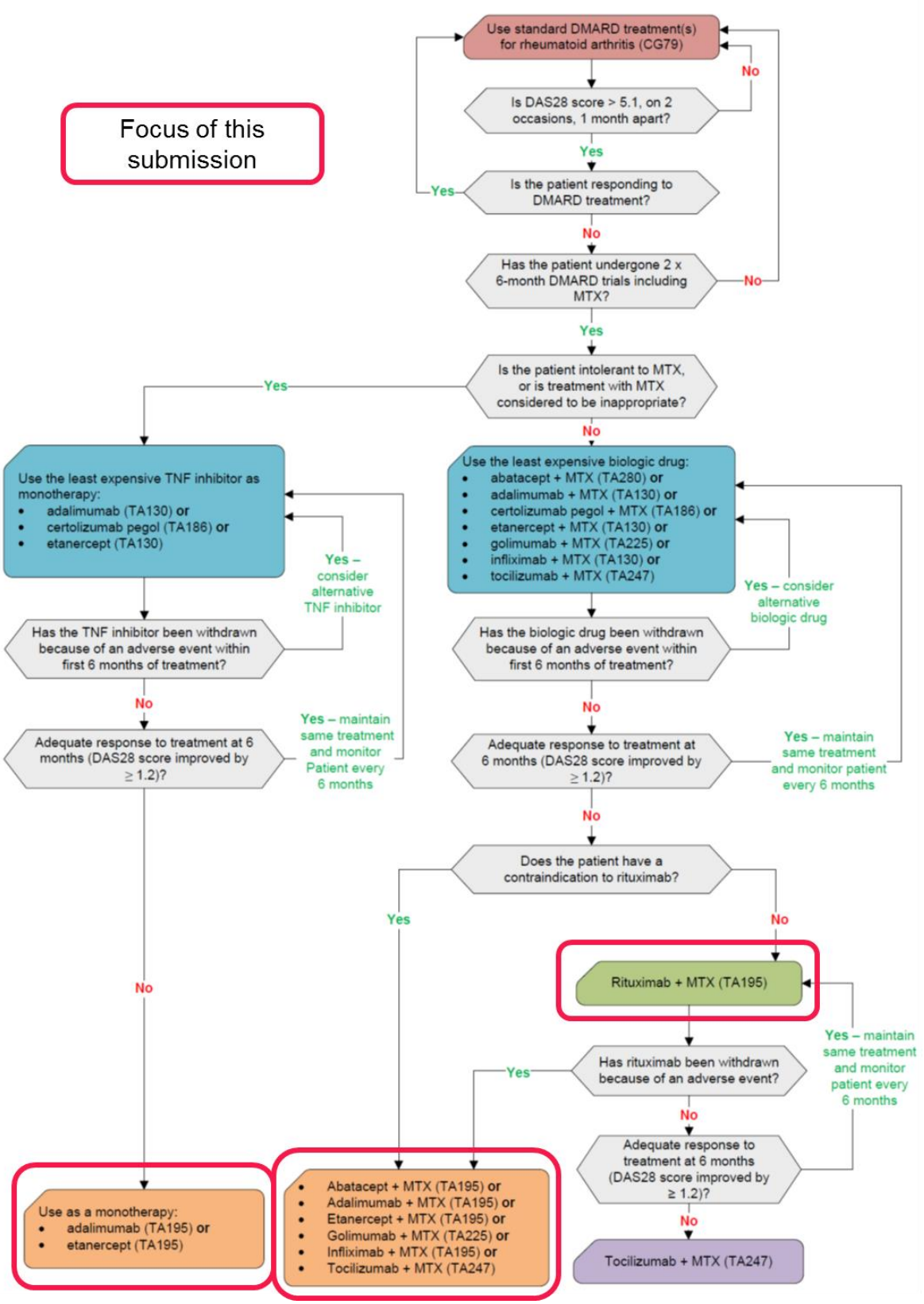
In the current NICE clinical pathway, CZP is only available as a first-line TNFi and has not yet been assessed in patients who have inadequate response to other first-line TNFi agents. Currently, all other TNFi therapies approved for use in RA in the UK in combination with MTX are recommend for use in inadequate responders to TNFi therapy, along with all currently approved monotherapy agents.[43, 44] In a recent MTA, the efficacy of CZP was shown to be similar to that of the other approved TNFi agents, therefore an approach to match the availability of CZP with the rest of the TNFi class is deemed appropriate.[14] This alignment is evident in the EULAR guidelines for the treatment of RA that state that no TNFi is better than any other in patients with active disease, despite initial treatment with a TNFi; therefore patients can access any of the available TNFi therapies after having had an inadequate response to an initial one.[46] It would thus be valuable to examine the potential

position of CZP within the NICE treatment pathway, to ensure alignment with the current EULAR treatment recommendations.

There are three positions within the current NICE clinical pathway that CZP can be considered as an alternative treatment option (Figure 2). Firstly, as an alternative to RTX plus MTX as a TNFi for patients who have failed to respond adequately to an initial TNFi. Secondly, CZP plus MTX may be added to the list of currently recommended TNFi agents for patients who had an inadequate response to their first TNFi but have a contraindication to RTX. Thirdly, CZP could also be added to the list of TNFi monotherapies for patients who have failed their first monotherapy and have a contraindication to MTX.

There are no equality issues arising in relation to this technology.

Figure 2: NICE clinical guidelines and technology position



Current treatment pathway for RA according to NICE guidelines. CZP can be considered into the pathway at any of the three highlighted stages: as a monotherapy following an inadequate response to a first TNFi, in combination with MTX following a first TNFi and a contraindication to RTX or in place of RTX and in combination with MTX following an inadequate response to a first TNFi.
Adapted from NICE RA commissioning algorithm.[63]

4 Clinical effectiveness

CZP is effective in patients with moderate to severe RA whose disease has not responded adequately to a prior TNFi.

- The efficacy of CZP in patients with moderate to severe RA previously exposed to TNFi was proven in six completed RCTs,[4-9] which have all provided consistent results and supported the efficacy of CZP in the TNFi-IR population, comparable with TNFi naïve populations.[4, 6]
- Both dosing regimens of CZP (CZP 200 mg Q2W and CZP 400 mg Q4W), compared to PBO, further provided clinically meaningful improvements to TNFi experienced patients, in improving sign and symptoms, reducing the disease activity, improving physical function, health-related quality of life, workplace and household productivity and participation in social activities.
- Within the largest study of TNFi experienced patients (n=400), examined as part of the REALISTIC study, treatment with CZP resulted in statistically significant improvements in both ACR20 and ACR50 response rates in comparison to PBO ($p<0.01$ and $p<0.05$, respectively).[4]

CZP has a rapid onset of action and short term response can predict long-term outcomes

- In the REALISTIC trial, ACR20 response rates were higher in the CZP group, compared to PBO, within the overall study population as early as the first assessment (Week 2, $p<0.001$).[4, 6]

CZP is similarly effective either as a monotherapy or in combination with MTX in patients with prior TNFi experience

- Overall treatment response by Week 12 was consistent across treatment groups, irrespective of prior or concomitant treatment, including prior TNFi use.[4, 6]
- Improvements to clinical and patient-reported domains, including ACR response, disease activity and fatigue, were shown to be comparable in TNFi experienced patients taking CZP as a monotherapy, or in combination with MTX.

Initial improvements with CZP treatment are maintained in the long term

- Improvements in disease activity, physical functioning, health-related quality of life and workplace and household productivity were maintained in the long-term up to Week 52.[6]
- Maintenance of benefit of CZP treatment for TNFi experienced patients was seen in both the CZP 200 mg Q2W and CZP 400 mg Q4W dosing arms, compared to PBO.[5, 8]

CZP has a favourable risk benefit profile in both TNFi naïve and TNFi experienced patients

- The AE profile of CZP was found to be in line with that previously reported for other TNFis approved for the treatment of RA in a long-term safety analysis, and no new safety signals emerged during the analysis of 4,049 CZP-treated patients.[10]
- Evidence for a comparable safety profile for CZP in the TNFi experienced versus overall study population was provided in the REALISTIC and SWITCH studies.[7, 64]
- A number of other safety assessments of RCTs including both patients naïve or previously exposed to another TNFi, namely DOSEFLEX and PREDICT, provided favourable safety data for the use of CZP.[5-7]
- Patients with prior TNFi exposure re-treated with CZP were recently shown to not be of higher risk of hospitalised infections compared to being re-treated with ABA, while risk of hospitalised infection was greater for patients subsequently re-treated with ETA, IFX or RTX.[12]

A systematic review was performed to assess the relative efficacy of CZP and its comparators listed in the scope for has not responded adequately to a prior TNFi. The indirect analysis conducted showed that CZP in combination with MTX is similar or more effective to the other comparators considered in all of the cases in patients with moderate to severe RA that were previously exposed to TNFi. The wide credible intervals noted in some analyses results reflect the minimal differences in relative clinical effect between CZP and the comparators considered. These results should be interpreted with caution due to heterogeneity in the patient population among the included studies.

4.1 *Identification and selection of relevant studies*

4.1.1 Systematic literature review

A systematic literature review was conducted to identify CZP trials in moderate to severe RA with an inadequate response to TNFi therapy. Six RCTs and one non-randomised observational study were identified that considered CZP in combination with MTX or cDMARDs, or as monotherapy, in the population of interest. The RCTs are summarised in Table 9. The evidence from RCTs is presented in Section 4.7 and the evidence for CZP from non-randomised trials is considered in Section 4.11.

4.1.2 Search strategy

Searches of the electronic databases and relevant conference proceedings were conducted up to 16th November 2015; relevant conference proceedings were searched for the last 4 years (2012–2015). The full search strategy is given in Appendix 8.2.

Further information on the methodology of the systematic review can be found in Section 4.1.3.

4.1.3 Study selection

To align with the NICE decision problem for the current STA, studies were screened to include those that recruited RA patients with an inadequate response to TNFi therapy, or those with a mixed population with a subgroup analysis of patients with an inadequate response to TNFi therapy.

An overview of the eligibility criteria is provided in Table 8. Further explanation and the rationale to justify these criteria are provided in the following sections.

Table 8: Eligibility criteria used in the search strategy

Clinical effectiveness	Inclusion criteria	Exclusion criteria
Population	Adults (≥18 years) with moderate to severe, active RA whose disease has not responded adequately to a TNFi	Paediatric RA patients; Adult RA patients naïve to TNFi
Intervention	• CZP monotherapy or in combination with MTX • RTX in combination with MTX • ABA, ADA, ETA, GOL, IFX and TOC each in combination with MTX • ADA monotherapy, ETA monotherapy or TOC monotherapy	Any other intervention
Comparators	• Any of the interventions above • Best supportive care/cDMARDs/PBO	Any other comparators

Outcomes	• Disease characteristics:
	  • ACR response rates (ACR20, ACR50, ACR70)
	  • DAS28(ESR) remission (DAS28(ESR) score <2.6)
		Any other outcomes

Clinical effectiveness	Inclusion criteria	Exclusion criteria
	• DAS28(ESR) low disease activity (DAS28(ESR) score <3.2) • EULAR response (none, good, moderate)	
Study design	• RCTs (triple/double/single blind or open label) • Non-RCT (only for CZP) • Comparative cohort studies/longitudinal studies (prospective or retrospective) (only for CZP) • Case-controlled study (only for CZP) • Comparative-cross-sectional study (only for CZP)	• Analysis of hospital records/database • Single arm studies (uncontrolled trials) • Case study • Case series • Case report • Systematic reviews
Language restrictions	English language only	Any other language

4.1.3.1 Study design

Further details of the study designs that were included and excluded are shown in Appendix 8.3.1.

RCTs were included regardless of blinding of the participants and the investigators. Open-label extensions of RCTs were also identified in the review.

Comparative non-randomised studies or comparative observational/real-world studies were searched for CZP only. Case series, case studies, and case reports were excluded, because they are generally smaller, non-comparative studies which carry a higher risk of bias compared to other study designs.

Relevant systematic reviews were highlighted during abstract screening, and bibliographic searching of the systematic reviews was conducted to identify relevant studies.

4.1.3.2 Population

The patient population of interest to the review comprised adult patients (≥18 years) of any race with moderate to severe, active RA whose disease has not responded adequately to a TNFi. Studies focussing solely on children and adolescents were excluded from this review. Studies enrolling a mixed patient population of children and adults were, however included if a subgroup analysis of the adult patients was provided.

4.1.3.3 Intervention

This systematic review focused on the bDMARDs including ABA, ADA, CZP, ETA, GOL, IFX, RTX, and TOC. These may be given as monotherapy or in combination with MTX. This systematic review focussed on the licensed doses of the interventions of interest.

The following points were considered for inclusion/exclusion:

- Studies including multiple interventions of interest should have at least one intervention with a licensed dose for inclusion in the review
- Studies in which the dose received by the patient population is unclear or studies in which the patients received both licensed and unlicensed doses with no subgroup analysis for licensed dose were excluded

4.1.3.4 Comparators

The study must have compared the intervention of interest to one of the comparators listed in Table 8.

4.1.3.5 Language

It is expected that the majority of high-quality evidence from RCTs in this field are available in the English-language literature. Therefore, non-English studies were not included in this review. Studies with an English abstract where the full-text is non-English were not included.

4.1.3.6 Publication timeframe

To ensure inclusion of all relevant evidence, database searches were conducted from January 1966 (database start) to 16th November 2015.

4.1.3.7 Data sources

The following databases were searched:

- MEDLINE (database start – November 2015)
- Embase (database start – November 2015)
- Cochrane Central Trials Register (database start – November 2015)
- MEDLINE-In process

MEDLINE and Embase were searched using the Embase.com interface. Cochrane Central Trials Register was searched using the Cochrane Library interface. MEDLINE-In process was searched using the Pubmed.com interface.

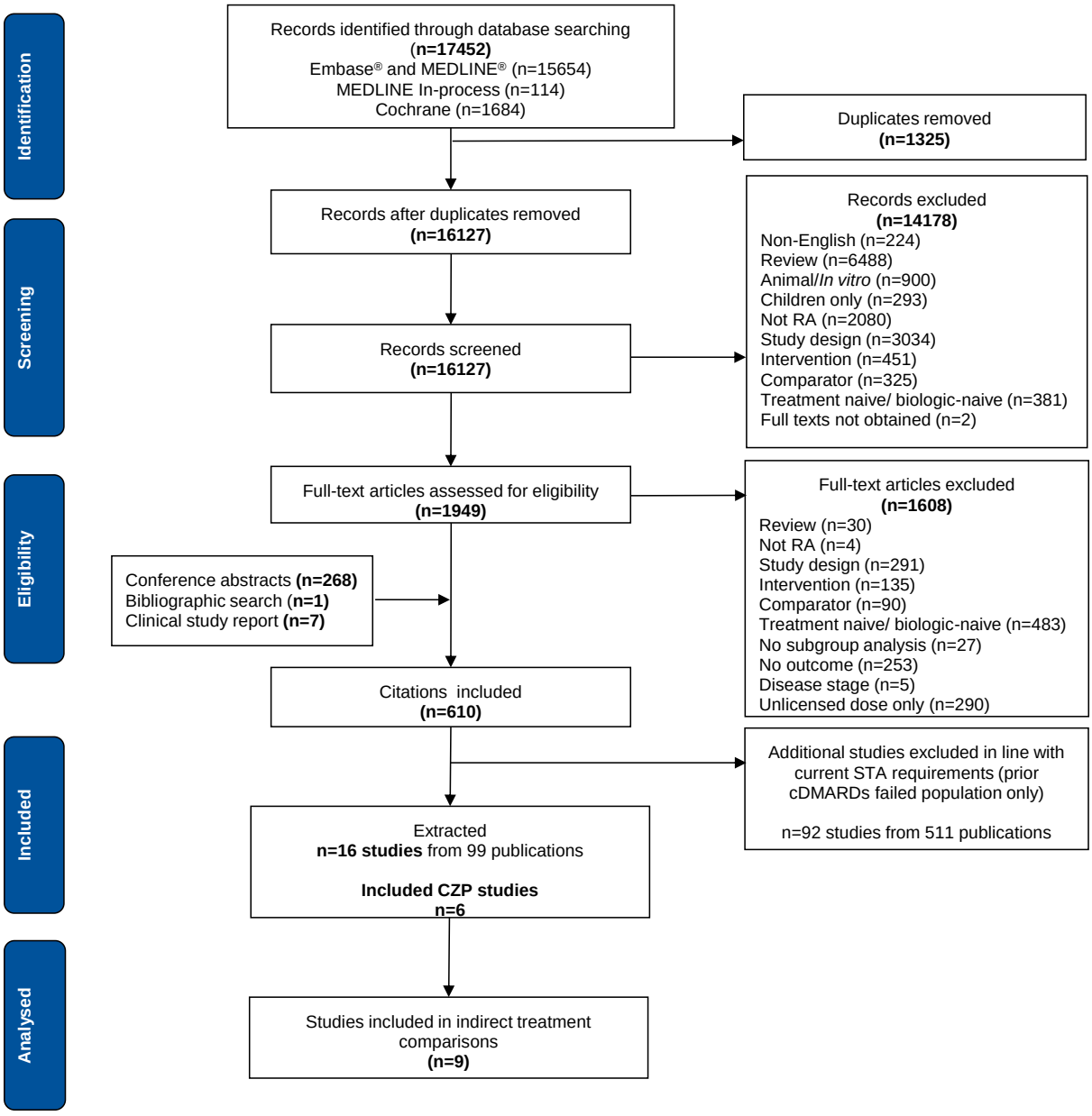
Conference abstracts were searched to retrieve the latest studies which have not yet been published in journals as full text articles or supplement results of previously published studies.

The ACR and EULAR conferences were searched for abstracts published between 2012 and 2015.

4.1.3.8 PRISMA flow diagram

A flow diagram of the number of studies included and excluded at each stage is presented in Figure 3.

Figure 3: PRISMA flow diagram



As shown in the PRISMA flow diagram, 16 studies (reported in 99 publications and 5 clinical study reports) met the inclusion/exclusion criteria of the systematic review.

Six of these studies provided data for CZP in patients with RA who have had inadequate response to a TNFi (REALISTIC, DOSEFLEX, SWITCH, J-RAPID and HIKARI).[4, 5, 7-9] It should be noted that the PREDICT RCT and ARTIS observational registry assessing CZP did not meet the full set of inclusion criteria of the systematic

review. The PREDICT study does not have a non-CZP control arm and thus did not meet the inclusion criteria for comparator.[6]

ABA, RTX and TOC were assessed in two studies each, while one study assessed GOL. Two small studies compared bDMARDs against each other; Combe 2012[65] compared RTX versus ETA, both in combination with MTX and OPPOSITE[62] study compared ETA versus IFX, both in combination with MTX. ROC was a pragmatic open-label study that compared the initiation of a second TNFi or of another class of bDMARD in patients with inadequate response to a first TNFi. The choice of the second TNFi (ADA, CZP, ETA, or IFX) or another bDMARD (ADA, RTX, or TOC) was at the discretion of the clinician.[66]

The systematic review did not identify any non-RCT or comparative observational study for certolizumab pegol. ARTIS study for certolizumab pegol was identified in the searches but was excluded because it is a single arm uncontrolled study.[67]

A full list of studies relevant to the decision problem is given in Table 9. A full list of studies included in the systematic review is given in Appendix 8.4.1. A list of studies that were included in the systematic review and were relevant to the decision problem but were excluded from the indirect treatment comparison, including the reason for exclusion, is given in Appendix 8.4.2.

4.2 *List of relevant randomised controlled trials*

4.2.1 Relevant RCTs

An overview of the included RCTs for CZP is provided in Table 9. All studies compare CZP with PBO, except PREDICT. It should be noted that for the purpose of the systematic literature review and meta-analysis, only licensed doses of the interventions of interest were considered ie. as monotherapy or in combination with MTX.

Table 9: List of relevant RCTs

Trial	Reference	Interventions	Patient population	Duration of study	Primary outcome(s)
Moderate to severe disease activity population					
REALISTIC (NCT00717236)	Weinblatt 2012[4]	• CZP 200 mg Q2W +/- MTX/cDMARDs* • PBO +/- MTX/cDMARDs	Active RA with an inadequate response to >1 prior cDMARD, having received treatment with ≤ 2 TNFis	28 weeks	• ACR20 response at 12 weeks
DOSEFLEX (NCT00580840)	Furst 2015[5]	• CZP 200 mg Q2W + MTX[≠] • CZP 400 mg Q4W + MTX[≠] • PBO + MTX[≠]	Active RA receiving MTX for ≥3 months, including patients with prior TNFi exposure	34 weeks[≠]	• ACR20 response at 34 weeks
PREDICT (NCT01255761)	Curtis 2015[6]	• CZP 200 mg Q2W +/- MTX/cDMARDs*	Active RA with unsatisfactory response or intolerance to ≥1 DMARD, having received treatment with ≤2 TNFis	52 weeks	• CDAI and RAPID-3 scores at 12 and 52 weeks • DAS28(ESR) at 52 weeks
SWITCH (NCT01147341)	Schiff 2014[7]	• CZP 200 mg Q2W + cDMARDs* • PBO + cDMARDs	Active RA having had inadequate response or intolerance to a TNFi other than CZP	24 weeks	• ACR20 response at 12 weeks
J-RAPID[§] (NCT00791999)	Yamamoto 2014[8]	• CZP 100 mg Q2W + MTX ** • CZP 200 mg Q2W + MTX * • CZP 400 mg Q2W + MTX * • PBO + MTX	Active RA with an inadequate response to MTX, including patients with prior exposure if they received 1 TNFi as a non-primary failure (only Japanese patients)	24 weeks	• ACR20 response at 12 weeks
HIKARI (NCT00791921)	Yamamoto 2014[9]	• CZP 200 mg Q2W +/- non-MTX cDMARDs[Ŧ] • PBO +/- non-MTX cDMARDs	Active RA with an inadequate response to ≥1 prior DMARDs (including MTX), including patients with prior exposure if they received 1 TNFi as a non-primary failure (only Japanese patients)	24 weeks	• ACR20 response at 12 weeks

*Patients received a loading dose of 400 mg CZP at Weeks 0, 2, and 4; **Patients received a loading dose of 200 mg CZP at Weeks 0, 2, and 4,; [≠] In DOSEFLEX, all patients received 400 mg CZP at Weeks 0, 2, and 4 followed by 200 mg CZP Q2W up to and including Week 16 in a run-in phase, followed by 16 weeks in randomisation to listed interventions.
[§]Trial contains CZP doses unlicensed in the European Union (CZP loading dose of 200 mg at Weeks 0, 2 and 4, for the CZP 100 mg Q2W dosage arm, and maintenance doses of CZP 100 mg Q2W and CZP 400 mg Q2W). [Ŧ]CZP in combination with non-MTX cDMARDs is not approved in the European Union.

The outcomes measured in the relevant RCTs which have been considered in this submission are given in Table 10.

Table 10: Overview of RCTs and outcomes considered

Trial	Outcome	Data source
REALISTIC (NCT00717236)	• ACR response • ACR component scores • EULAR response • DAS28 (ESR and CRP) score • DAS28(ESR) remission • Fatigue and sleep • Physical function (HAQ-DI) • CDAI	• Weinblatt 2012 for ACR response rates at Week 12 • UCB DoF tables covering ACR response, ACR component scores, EULAR response, DAS28 scores, DAS28 remission, fatigue and sleep, HAQ-DI and CDAI for all other time points and OLE
DOSEFLEX (NCT00580840)	• ACR response • ACR component scores • EULAR response • DAS28(ESR) score • DAS28(ESR) remission • Fatigue • Physical Function (HAQ-DI) • CDAI • SF-36 (PCS and MCS)	• Furst 2015 for ACR response at Week 34 • UCB DoF tables covering ACR response, ACR component scores, EULAR response, DAS28 scores, DAS28 remission, HAQ-DI, CDAI and SF-36
PREDICT (NCT01255761)	• mACR response • mACR component scores • EULAR response • DAS28(ESR) score • DAS28(ESR) remission • Multidimensional-HAQ (MD-HAQ) • CDAI • Workplace and household productivity (WPS-RA)	• UCB DoF tables covering mACR response, mACR component scores, EULAR response, DAS28 scores, DAS28 remission, MD-HAQ, CDAI and WPS-RA
SWITCH (NCT01147341)	• ACR response • DAS28 (CRP) score • Physical function (HAQ-DI)	• Schiff 2014 for all outcomes
J-RAPID (NCT00791999)	• ACR response • EULAR response • DAS28(ESR) remission	• UCB DoF for all outcomes
HIKARI (NCT00791921)	• ACR response • EULAR response • DAS28(ESR) remission	• UCB DoF for all outcomes

CRP: C-reactive protein; DoF: Data on File; MCS: mental component score; PCS: physical component score.

4.2.2 Exclusion of RCTs from further discussion

None of the included RCTs are to be excluded from further discussion.

4.3 *Summary of methodology of the relevant randomised controlled trials*

The trial information and study methodology used in the six relevant RCTs are presented below. Five of the RCTs were PBO-controlled studies, whilst one (PREDICT) treated all patients with CZP 200 mg Q2W in combination with MTX/cDMARDs.

The PREDICT trial examined the predictability of treatment success 12 weeks after starting therapy in moderate to severe RA patients receiving CZP, using both the patient-reported RAPID3 tool and the investigator-based CDAI assessment tool. The study consisted of a 52 week double-blind phase between patients randomised by the different assessment tools.[6]

The five other RCTs had different designs; REALISTIC and SWITCH both commenced with a 12 week double-blind phase for patients treated with either CZP or PBO, followed by OLE periods.[4, 7] DOSEFLEX commenced with a run-in phase up to Week 16 whereby all patients were treated with CZP in combination with MTX; from Week 18 onwards, patients were randomised to CZP or PBO, each in combination with MTX, for a 16 week double-blind phase.[5] DOSEFLEX was followed by DOSEFLEX II (NCT00753454), an OLE phase enrolling patients who completed the 34 week DOSEFLEX study. Data from DOSEFLEX II is not presented in this submission. The two Japanese trials, J-RAPID and HIKARI, both entailed a 24 week, double-blind, randomised phase comparing CZP treatment groups with PBO, with or without concomitant cDMARDs.[8, 9] J-RAPID patients were treated with either CZP or PBO in combination with MTX, while participants in the HIKARI study received CZP or PBO with or without non-MTX DMARDs.[*] Both trials were followed by OLEs, but data from these extension periods are not presented as part of this submission.

A comparison of the methodology of the RCTs eligible for inclusion is presented in Table 11 and Table 12.

[*]CZP in combination with non-MTX cDMARDs is not approved in the European Union.

Table 11: Comparative summary of trial methodology

	REALISTIC (NCT00717236)		DOSEFLEX (NCT00580840)		PREDICT (NCT01255761)		SWITCH (NCT01147341)	
Location	230 centres in 7 countries in the USA and Canada (75%) and Europe (25%)		63 centres in 3 countries		110 centres in 1 country		12 centres in 1 country	
Trial design	Randomised, double-blind, PBO-controlled, parallel group		Randomised, double-blind, PBO-controlled, parallel group, open-label run-in		Randomised, double-blind, parallel group		Randomised, double-blind, PBO-controlled, parallel group	
Eligibility criteria for participants	Inclusion: • Aged ≥18 yrs • Adult onset RA (ACR definition) for at least 3 months • Active RA at screening and baseline, defined as: - ≥5 tender joints (28 joint count) AND - ≥4 swollen joints (28 joint count) AND - At least one of the following two criteria: - CRP >10 mg/L - ≥28 mm/h ESR (Westergren) • Failure to respond (lack of efficacy or intolerance) to at least one synthetic DMARD • Discontinuation of all ineligible DMARD therapy at least 28 days or 5 half-lives prior to first dose of study drug	Exclusion: • A diagnosis of any other inflammatory arthritis, e.g. PsA or AS • History of chronic, serious or life-threatening infection, • Any current infection • Active or a history of active tuberculosis, • Receipt of a biologic therapy for RA within 2 months (1 month for ETA or anakinra) prior to baseline • Prior receipt of either >2 TNFis or RTX and/or ABA or discontinuation of a biologic therapy for RA due to severe hypersensitivity or anaphylactic reactions	Inclusion: • Aged ≥18 yrs • Postmenopausal for at least 1 year (females) or incapable of childbearing • Diagnosis of adult-onset RA, having lasted 6 months-15 years (ACR 1987 criteria) • Rheumatoid factor positive and/or anti-CCP positive • Active RA at screening and baseline, defined as: - ≥6 tender joints (28 joint count) AND - ≥4 swollen joints (28 joint count) AND - At least one of the following: - CRP >10 mg/L - ≥28 mm/h ESR (Westergren) • Received treatment with MTX (10 to 25 mg/week, with or	Exclusion: • Inflammatory arthritis (or secondary, non-inflammatory arthritis) other than RA • History of an injected joint prosthesis at any time with that prosthesis still in situ • Use of prohibited medication (analgesics, oral corticosteroids, IA hyaluronic acid, specific DMARDs) • Receiving experimental therapy for RA (or other conditions) • Failure to respond to previous treatment with TNFis (primary failures) • Chronic infection, recent serious or life-threatening	Inclusion: • Aged ≥18 years • A diagnosis of adult-onset RA of ≥3 months duration at baseline (defined as 1987 ACR classification criteria): • ≥4 tender joints (28 joint count) • ≥4 swollen joints (28 joint count) • Unsatisfactory response or intolerance to ≥1 DMARD • Treatment with ≤2 TNFis prior to enrolment	Exclusion: • A diagnosis of any other inflammatory arthritis • A diagnosis of secondary, non-inflammatory type of arthritis, symptomatic enough to interfere with evaluation of the effect of CZP • A diagnosis of fibromyalgia with sufficient symptoms requiring treatment	Inclusion: • RA diagnosed by the 1987 ACR criteria of >6 months duration, functional Class 1 – 3. • Active RA at screening and baseline, defined as: - ≥6 tender joints (28 joint count) AND - CRP >10 mg/L - CDAI ≥12 at screening • Previous secondary inadequate response or intolerant of a TNFi other than CZP • Oral MTX or other cDMARD continuously for ≥3 months before first study dose	Exclusion: • B cell depleting agent taken within the 6 months prior to enrolment

	(whichever was longer)		without folic acid) for at least 3 months prior to the baseline visit	infection, or current sign of infection • High risk of infection • History of active or latent tuberculosis • History of lymphoprolif-erative disorder • Any vaccination within 8 weeks prior to baseline				
Settings and locations where the data were collected	Study site visits		Study site visits		Study site visits		Study site visits	
Intervention and comparator(s)	• CZP 200 mg Q2W +/- MTX/cDMARDs* (n=851) • PBO +/- MTX/cDMARDs (n=212)		• CZP 200 mg Q2W + MTX† (n=70) • CZP 400 mg Q4W + MTX† (n=70) • PBO + MTX† (n=69)		• CZP 200 mg Q2W +/- MTX/cDMARDs* - RAPID3 (n=368) - CDAI (n=365)		• CZP 200 mg Q2W + cDMARDs* (n=27) • PBO + cDMARDs (n=10)	
Primary outcomes (including scoring methods and timings of assessments)	• ACR20 response at 12 weeks		• ACR20 response at 34 weeks		• CDAI and RAPID-3 scores at 12 and 52 weeks • DAS28(ESR) at 52 weeks		• ACR20 response at 12 weeks	
Secondary/ tertiary outcomes (including scoring methods and timings of assessments)	• ACR20 response at Week 12 based on pre-specified baseline stratification factors (see subgroups below)) • ACR20/50/70 response at Weeks 12 and 28 • Improvements in ACR component scores and EULAR response at Weeks 12 and 28 • DAS28 reduction based on DAS28 (CRP), SDAI and CDAI • DAS28(ESR) remission rates at Weeks 12 and 28 • TJC, SJC, HAQ-DI, CRP improvement • Improvement in pain, fatigue and sleep problems • Time to sustained ACR20 response and EULAR response • Safety		• ACR20/50/70 responses at Weeks 16 and 34 • ACR component scores at Weeks 16 and 34 • EULAR response rates at Weeks 16 and 34 • DAS28(ESR) response, remission rates and HAQ-DI up to Week 34 • CDAI, SDAI and DAS28(ESR) remission at Weeks 16 and 34 • PROs (SF-36 component summaries and domain scores, fatigue and pain) at Week 34 • Safety		• mACR20/50/70 response rates • mACR component scores and EULAR response rates at Weeks 12, 24 and 52 • DAS28(ESR) response and remission rates by disease activity up to Week 52 • CDAI, RAPID3 and DAS28(ESR) remission up to Week 52 • MD-HAQ, and work and household productivity improvement up to Week 52 • Safety		• CDAI response, ACR50/70 response, LDA (DAS28 (CRP) of ≤3.2 or CDAI <10) at Weeks 12 and 24 • Change from baseline in HAQ-DI • Safety	

Pre-planned subgroups	Subjects were stratified by: • Baseline MTX use • Previous TNFi use • Disease duration (<2 vs ≥2 years)	Analyses by prior TNFi therapy at baseline up to Week 34 were not planned in the protocol but performed post-hoc.	None	None.

CCP: cyclic citrullinated peptide; SDAI: simple disease activity index; TJC: tender joint count; SJC: swollen joint count.
*Patients received a loading dose of 400 mg CZP at Weeks 0, 2, and 4; †In DOSEFLEX, all patients received 400 mg CZP at Weeks 0, 2, and 4 followed by CZP 200 mg Q2W up to Week 18 in a run-in phase, followed by 16 weeks in randomisation – the patient numbers presented for the DOSELFEX trial refer to the period of randomisation.

Table 12: Comparative summary of trial methodology (continued)

	J-RAPID (NCT00791999)		**HIKARI (NCT00791921)**	
Location	Japan		Japan	
Trial design	Randomised, double-blind, placebo-controlled, parallel group		Randomised, double-blind, placebo-controlled, parallel group	
Eligibility criteria for participants	Inclusion: • Aged 20-75 yrs • Diagnosis of adult-onset RA of 0.5−15 years' duration as defined by ACR (1987) criteria • Active RA at screening and baseline, defined as: - ≥9 tender joints (68 joint count) AND - ≥9 swollen joints (66 joint count) - At least one of the following: - ESR ≥30 mm/hour - CRP ≥1.5 mg/dL • Received treatment with MTX for ≥6 months with fixed dose for prior ≥2 months at 6-8 mg/week	Exclusion: • Treatment with any biologic therapy for RA within the 6 months preceding the study (3 months for ETA) • Any investigational drug in the preceding 3 months • Previous treatment with ≥2 TNFis; or previously failed to respond to TNFi therapy • Previous severe hypersensitivity or anaphylactic reaction following TNFi treatment	Inclusion: • Aged 20-75 yrs • Diagnosis of adult-onset RA of 0.5-15 years' duration as defined by ACR (1987) criteria • Failed treatment with or resistant to ≥1 prior DMARD (including MTX), or unable to receive MTX because of safety concerns • Active RA at screening and baseline, defined as: - ≥6 tender joints (68 joint count) AND - ≥6 swollen joints (66 joint count) - At least one of the following: - ESR ≥28 mm/hour - CRP ≥2.0 mg/dL • Non-MTX DMARDs permitted if fixed dose for prior ≥2 months	Exclusion: • Inflammatory arthritis other than RA • Any biologic treatment for RA in the 6 months preceding the study (3 months for ETA) • Any investigational drug in the preceding 3 months • Previous treatment with ≥2 TNFis, or previously failed to respond to TNFi therapy • Previous severe hypersensitivity or anaphylactic reaction following bDMARDs • Azathioprine and cyclosporine in 28 days prior to start of trial drug administration
Settings and locations where the data were collected	Study site visits		Study site visits	

Intervention and comparator(s)	• CZP 100 mg Q2W + MTX**† (n=72) • CZP 200 mg Q2W + MTX* (n=82) • CZP 400 mg Q2W + MTX*† (n=85) • PBO + MTX (n=77)	• CZP 200 mg Q2W +/- non-MTX cDMARDs‡ (n=116) • PBO +/- non-MTX cDMARDs (n=114)
Primary outcomes (including scoring methods and timings of assessments)	• ACR20 response at 12 weeks	• ACR20 response at 12 weeks
Secondary/ tertiary outcomes (including scoring methods and timings of assessments)	• ACR20 response at 24 weeks • ACR50/70 response rates • ACR core component scores incl. HAQ-DI, pain, CRP and ESR • mTSS at Week 24 • DAS28(ESR) • EULAR response • SF-36 at Week 12/24	• ACR20 response at 24 weeks • ACR50/70 response rates • ACR core component scores incl. HAQ-DI, pain, CRP and ESR • mTSS at Week 24 • DAS28(ESR) • EULAR response • SF-36 at Week 12/ 24
Pre-planned subgroups	None	Post-hoc analyses of • CZP monotherapy • CZP with concomitant non-MTX cDMARDs

* Patients received a loading dose of 400 mg CZP at Weeks 0, 2, and 4;
** Patients received a loading dose of 200 mg CZP at Weeks 0, 2, and 4, which is not an approved loading dose in the European Union
† For maintenance in RA, CZP 400 mg Q2W and CZP 100 mg Q2W are not approved doses in the European Union
‡CZP in combination with non-MTX cDMARDs is not approved in the European Union.

4.3.1 REALISTIC

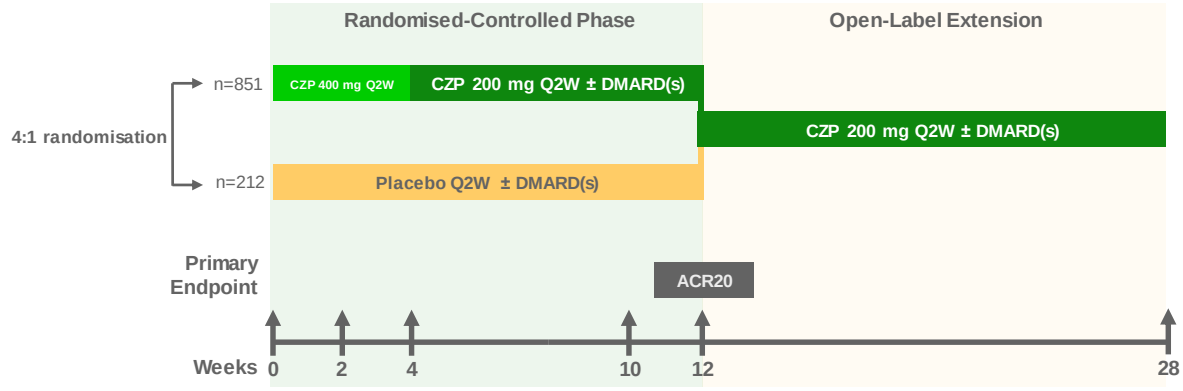
The REALISTIC trial (n=1,063) was a Phase IIIb, 12-week, randomised, double-blind, PBO-controlled study conducted between July 2008 and March 2010 in 230 centres in the USA, Canada and Europe. The 12-week, double-blind, PBO-controlled phase of the trial was followed by an OLE for at least 16 weeks. The REALISTIC study was designed to investigate the benefits of CZP in a broader, clinically relevant population of RA patients, including those with prior TNFi use and a range of concomitant DMARDs, including monotherapy, more closely resembling those seen in daily clinical practice.[4]

Analgesics, oral corticosteroids (≤10 mg/day prednisone equivalent), and NSAIDs/COX-2 inhibitors were permitted if the doses were stable within 24 hours, 7 days and 14 days of baseline, respectively. Intra-articular and parenteral corticosteroids, intra-articular hyaluronic acid within 4 weeks of baseline and DMARDs such as cyclosporine, cyclophosphamide, mycophenylate mofetil, chlorambucil and penicillamine within 3 months of baseline were prohibited. Patients were allowed to use the following DMARDs at the same stable dosing as used at baseline, until Week 12: MTX, leflunomide, sulfasalazine, chloroquine or hydroxychloroquine, azathioprine and gold.[4] Patient subgroups included CZP monotherapy, CZP in combination with MTX with or without further concomitant cDMARDs as listed above, and CZP in combination with concomitant cDMARDS.

As per study protocol, patients with prior TNFi experience were permitted to participate in the trial. ETA and anakinra should have been discontinued at least 1 month before study entry, and other biologic RA therapies within 2 months of study entry. Patients were excluded who received treatment either with more than two TNFis, RTX or ABA.[4]

In the 12-week double-blind phase of the trial, patients were randomised 4:1 and stratified by baseline MTX use, prior TNFi use, and disease duration (<2 years vs. ≥2 years), to receive either CZP (administered as the loading dose of CZP 400 mg at Weeks 0, 2 and 4, followed by a maintenance dose of CZP 200 mg Q2W) or PBO injection (control) Q2W in addition to their current treatment (if any) allowed by the treatment protocol described above. Patients who completed the 12-week double-blind phase (in both the CZP and PBO arms) were eligible to receive open-label CZP 200 mg Q2W for ≥16 weeks. For the OLE data reported, only data from those patients who were randomised to CZP in Week 0 will be presented in this submission. The OLE required a minimum of 16 weeks open label treatment and 12 weeks of safety follow-up. A schematic of the trial design is displayed in Figure 4.[4]

Figure 4: REALISTIC trial design schematic



The primary efficacy end-point was ACR20 response at Week 12. Secondary efficacy end-points assessed at Week 12 included:

- ACR50 (50% improvement in ACR score) and ACR70 (70% improvement in ACR score) responses
- Reduction of disease activity by DAS28 joint assessment based on ESR or CRP, in addition to SDAI and CDAI scores
- Improvement in the individual components of the ACR criteria, including TJC, SJC, HAQ-DI, CRP, patient's assessment of arthritis pain, and patient's and physician's global assessment of disease activity

Additional secondary end-points were time to sustained ACR20 response, which was defined as the time from randomisation to sustained ACR20 response at 2 consecutive visits (at the latest on Week 12) and EULAR response.[4]

Patient-reported outcomes represent an important measure of the burden of RA, and have a significant impact on the quality of the patients' life. In the REALISTIC study, fatigue (fatigue assessment scale) and sleep quantity and quality (sleep problem index [SPI]-II domain of the Medical Outcomes Study sleep scale) were also assessed. The minimal clinically important difference (MCID) was ≥1 point improvement for patient assessment of fatigue and ≥6 point improvement for the SPI-II.[68]

Efficacy and safety evaluations were assessed at baseline and at Weeks 2 (first post-baseline assessment), 6, 12 and every 8 weeks in the OLE and at the completion/withdrawal visit. Additional safety parameters were assessed at the follow-up visit (12 weeks after the last dose).[4]

Additional information related to participant eligibility, study locations and secondary objectives of the trial can be found in Table 11.

4.3.2 DOSEFLEX

DOSEFLEX was a 34-week, Phase IIIb, multicentre, double-blind, randomised, open-label run-in PBO-controlled study in patients with active moderate to severe RA and an incomplete response to MTX, conducted across the USA, Canada and France. The study was designed to compare the clinical efficacy and safety of two maintenance dosing regimens of CZP 200 mg Q2W and 400 mg Q4W vs. PBO, in combination with MTX, after an initial 16-week open-label run-in phase of CZP 200 mg Q2W in combination with MTX.[5]

During the open-label run-in phase all patients received the CZP loading dose of 400 mg CZP at Weeks 0, 2 and 4, followed by a maintenance dose of CZP 200 mg Q2W up to Week 16 in combination with MTX. Patients who had achieved an ACR20 response at Week 16 were randomised at Week 18 1:1:1 to receive PBO, or one of either CZP maintenance doses (200 mg Q2W or 400 mg Q4W), in combination with MTX, for a further 16 weeks up to Week 34.[5] [redacted]

Patients who failed to initially respond to previous TNFi treatment (primary non-responders) were excluded. TNFi responders who later discontinued that drug due to loss of efficacy or other reasons were eligible, provided that previous biologic therapy was stopped 3 months before baseline, except for ETA or anakinra (1 month).[5]

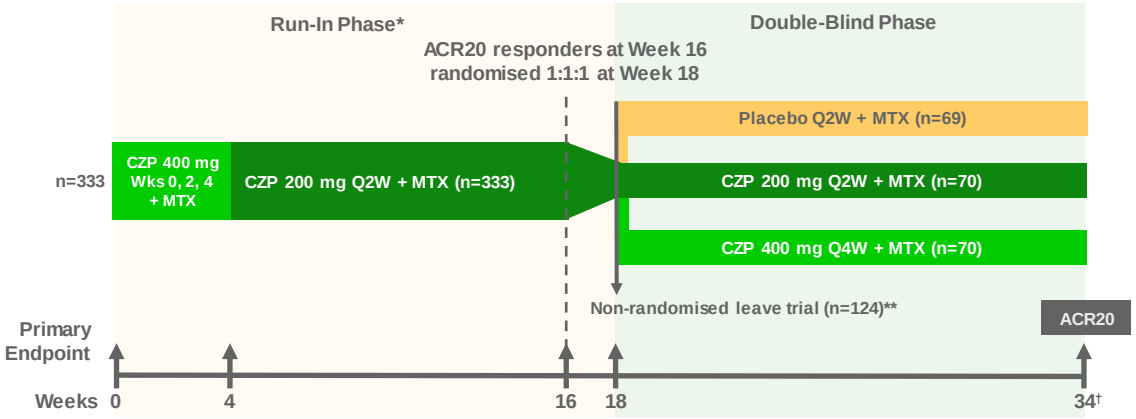
Concomitant treatment was allowed with analgesics, NSAIDs/COX-2 inhibitors and corticosteroids (prednisone or equivalent, 10 mg/day). Corticosteroid doses could be reduced according to local guidelines; dose increases were not permitted.[5]

The primary efficacy end-point was ACR20 response at Week 34. Secondary end-points included:[5]

- ACR20, ACR50 and ACR70 responses at Weeks 16 and 34
- DAS28(ESR) response and HAQ-DI up to Week 34
- CDAI, SDAI and DAS28(ESR) remission at Weeks 16 and 34
- Patient-reported outcomes (SF-36 component summaries and domain scores, fatigue and pain) at Week 34

The study design of DOSEFLEX is presented in Figure 5.

Figure 5: DOSEFLEX trial design schematic



*Last study medication dose during the run-in phase was at Week 16
**Patients who did not respond adequately at Week 16 were withdrawn at Week 18
†Randomised patients who completed the Week 34 assessment or experienced a flare during the double-blind phase were eligible for entry into an OLE (DOSEFLEX II)
For maintenance in RA, CZP 400 mg Q4W before clinical response is confirmed is not an approved dose in the European Union.

4.3.3 PREDICT

PREDICT was a 52-week, Phase IV, multicentre, double-blind, randomised, PBO-controlled study in patients with active moderate to severe disease activity as defined by the 1987 ACR criteria, from 110 centres in the USA. The study was designed to compare the sensitivity and predictive value of the RAPID-3 versus the CDAI tools in predicting treatment response at 1 year using Week 12 data.[6]

During this double-blind trial patients were randomised 1:1 to be monitored either by RAPID-3 or CDAI. All patients received the CZP loading dose (400 mg CZP at Weeks 0, 2 and 4), followed by a maintenance dose of CZP 200 mg Q2W up to Week 52. Until Week 12, stable treatment with the following concomitant cDMARDs was permitted if already used at baseline: MTX, leflunomide, sulfasalazine, chloroquine or hydroxycholorquine, antimalarials, and gold. At Week 12, patients who were classified as either RAPID-3 or CDAI failures were withdrawn. After Week 12, any patients classified as experiencing high disease activity at more than 2 consecutive visits were also withdrawn.[6]

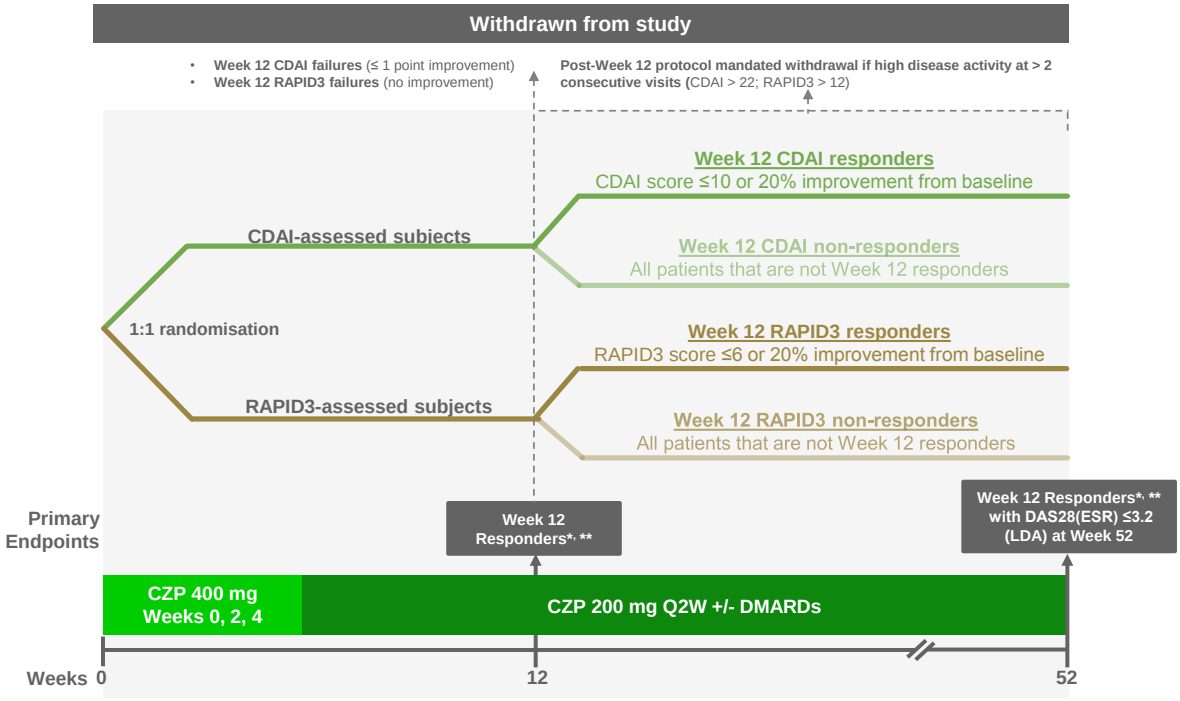
Patients with prior TNFi experience were permitted to participate in the trial. Patients were excluded if they had received treatment with ≥3 TNFi agents or any non-anti-TNF biologic agent prior to enrolment.[6]

The primary efficacy end-points were Week 12 predicted response as assessed by either RAPID-3 (≥20% improvement in baseline score or score of ≤6 points) or CDAI (≥20% improvement in baseline score or score of ≤10 points) and Week 12 responders as assessed by either RAPID-3 or CDAI achieving low disease activity (DAS28(ESR) ≤3.2) at Week 52. Secondary end-points included:

- Change from baseline at Weeks 12 and 52 in DAS28(ESR) score and disease activity status
- Change from baseline and disease status by CDAI at Weeks 12 and 52
- Change from baseline and disease status by RAPID-3 at Weeks 12 and 52

The study design of PREDICT is presented in Figure 6.

Figure 6: PREDICT trial design schematic



*CDAI responders are defined as subjects with a CDAI score ≤10 or 20% improvement from baseline
**RAPID3 responders are defined as subjects with a RAPID3 score ≤6 or 20% improvement from baseline

4.3.4 SWITCH

SWITCH was an investigator-initiated 12-week, Phase IV, multicentre, double-blind, randomised, PBO-controlled study in patients with active RA, as diagnosed by the ACR 1987 criteria. The study was designed to investigate the effect of CZP or PBO on measures of disease activity in patients with RA who had discontinued a TNFi other than CZP for secondary lack of efficacy or lack of tolerability.[7]

At the start of the 12 week double-blind phase, patients were randomised 2:1 to either CZP (administered as the loading dose of 400 mg at Weeks 0, 2 and 4, followed by a maintenance dose of CZP 200 mg Q2W) or PBO, in combination with cDMARDs.[7]

Patients must have taken oral MTX or other cDMARDs continuously for at least 3 months before the first study dose. Oral corticosteroids (≤10 mg/day prednisone equivalent) were permitted with stable dosing within 1 month of baseline and throughout the study period.[7]

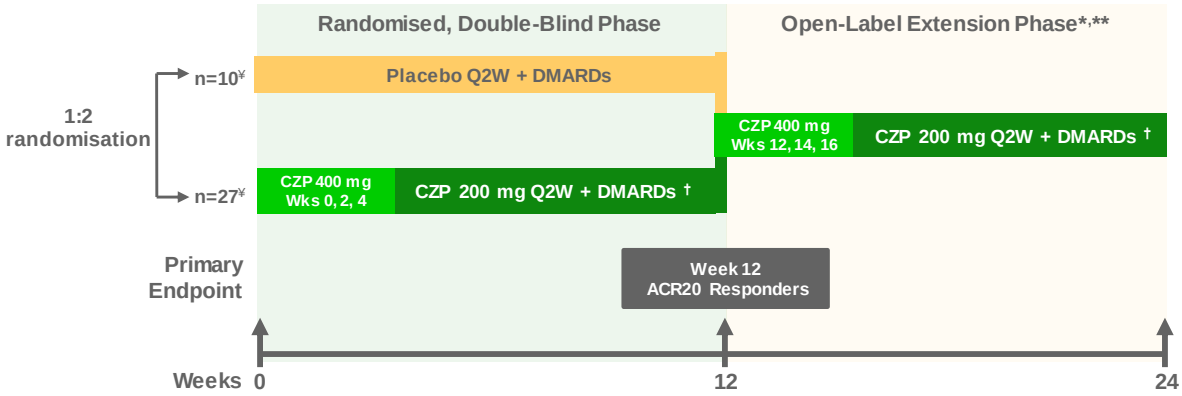
Patients had previous secondary inadequate response or were intolerant to a TNFi other than CZP, which was discontinued at least 28 days prior to the baseline visit. Of the 37 participants recruited, 34 patients had loss of efficacy, and 3 patients were intolerant (migraine [n=1], injection site reaction [n=1], infusion reaction and injection site reaction [n=1]) to another TNFi.[7]

The primary efficacy end-point was ACR20 response at Week 12. Secondary endpoints were also measured at Week 12 and included:[7]

- The proportion of patients achieving a CDAI response
- The proportion of ACR50 and ACR70 responders
- Low disease activity score (DAS28(CRP) of ≤3.2 or CDAI of <10)
- The mean change from baseline in HAQ-DI

The study design of SWITCH is presented in (Figure 7).

Figure 7: SWITCH trial design schematic



*To maintain blinding, the loading dose of CZP 400 mg at Weeks 12, 14 and 16 was given to all patients
**All patients completed the 12 week study, though one patient initially treated with PBO dropped out of the OLE
¥After randomisation of 37 (36.3%) of the 102 originally planned patients, enrolment into the study was stopped by the sponsor investigator, since an interim analysis demonstrated that the endpoint hypothesis was met

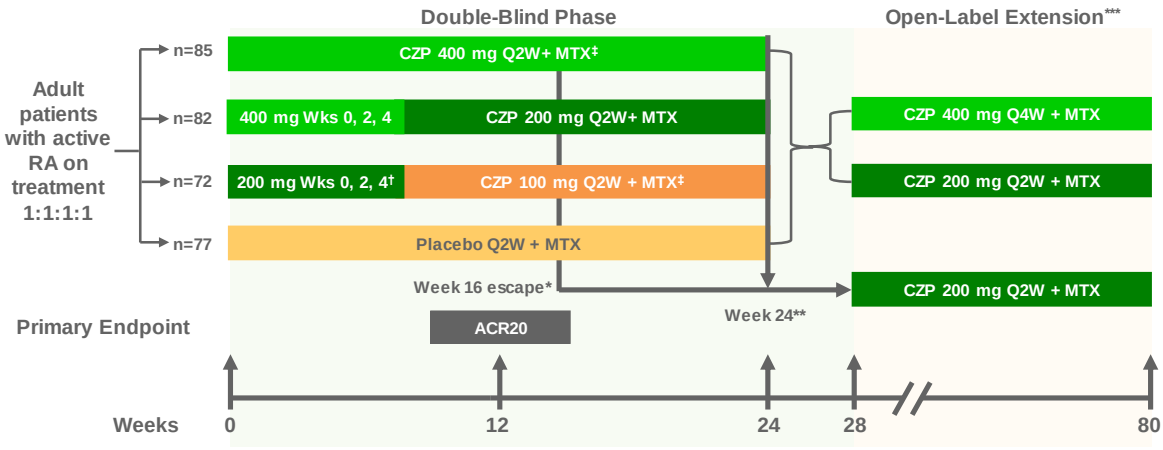
4.3.5 Other relevant RCTs

4.3.5.1 J-RAPID

J-RAPID is a 24-week, Phase II/III, multicentre, double-blind, randomised, PBO-controlled study in which Japanese patients with active RA and an inadequate response to MTX were randomised (1:1:1:1) to 1 of 4 treatment groups: CZP (100 mg Q2W, 200 mg Q2W, or 400 mg Q2W) or PBO, each in combination with MTX.[8] The objective of this trial was to investigate the efficacy of three dose regimens of CZP vs. PBO, each in combination with MTX in active RA patients who had an incomplete response to MTX.[8] Patients in the CZP groups received induction dosing with CZP 200 mg (CZP 100 mg Q2W group) or CZP 400 mg (CZP 200 mg Q2W and 400 mg Q2W groups) at Weeks 0, 2 and 4.

Patients who did not achieve an ACR20 response at Weeks 12 and 14 were withdrawn from the study at Week 16 and were eligible to enter an OLE study receiving CZP 200 mg Q2W, as were patients achieving an ACR20 response at Weeks 12 and 14 but failing to do so at Week 24.[69] Patients completing the 24 weeks of study as ACR20 responders were eligible to join the OLE, but were randomised 1:1 to the CZP 400 mg Q4W or CZP 200 mg Q2W treatment arm, both in combination with MTX treatment. The primary efficacy end-point was ACR20 response at Week 12.[69] Figure 8 describes the study design of J-RAPID.

Figure 8: J-RAPID trial design schematic

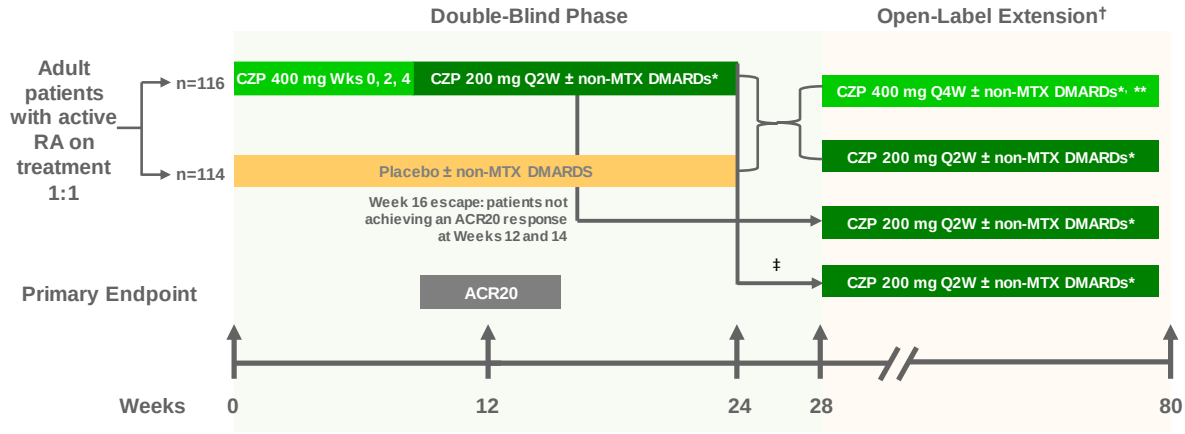


*Patients who did not achieve an ACR20 response at Weeks 12 and 14 were withdrawn from the study and were eligible to enter an OLE study thereafter and were treated with CZP 200 mg Q2W.
**Patients exhibiting ACR20 response at Weeks 12 or 14 but failing to achieve ACR20 response at Week 24 received CZP 200 mg Q2W in the OLE study.
***Patients who achieved an ACR20 response at Weeks 12 or 14 as well as at Week 24 were randomised (1:1) to either CZP 200 mg Q2W or CZP 400 mg Q4W in the OLE. For maintenance in RA, CZP 400 mg Q4W before clinical response is confirmed is not an approved dose in the European Union.
†A loading dose of CZP 200 mg at Weeks 0, 2 and 4 is not approved in the European Union.
‡For maintenance in RA, CZP 400 mg Q2W and 100 mg Q2W are not approved doses in the European Union.
Figure adapted from Yamamoto 2014[8], Takeuchi 2015[70], Takeuchi 2012[71]

4.3.5.2 HIKARI

HIKARI was a 24 week, Phase III, double-blind, randomised, PBO-controlled study in which patients were randomised to CZP 200 mg Q2W (following the CZP loading dose of 400 mg at Weeks 0, 2 and 4) or PBO, in combination with cDMARDs excluding MTX. Like J-RAPID, this trial was part of a development programme initiated in view of obtaining a regulatory approval of CZP for use in eligible RA patients in Japan. Patients not achieving ACR20 at Weeks 12 and 14 withdrew at Week 16 and were eligible to enter an OLE at CZP 200 mg Q2W. Patients finishing the 24-week double-blind phase as ACR20 responders were eligible for treatment with CZP 400 mg Q4W or CZP 200 mg Q2W, after randomisation. The primary efficacy end-point was ACR20 response at Week 12.[72] Figure 9 describes the study design of the HIKARI trial.

Figure 9: HIKARI trial design schematic



*CZP is indicated in combination with MTX in the European Union. CZP may be given as monotherapy in case of intolerance to MTX, or when continued treatment with MTX is inappropriate. CZP in combination with non-MTX cDMARDs is not approved in the European Union.
**For maintenance in RA, CZP 400 mg Q4W before clinical response is confirmed is not an approved dose in the European Union.
†CZP treatment without the loading dose is not approved in the European Union.
‡Patients exhibiting ACR20 response at Weeks 12 or 14 but failing to achieve ACR20 response at Week 24 were assigned to CZP 200 mg Q2W in the OLE study.
Figure adapted from: Yamamoto 2014[9], Takeuchi 2015[70], Takeuchi 2012[71]

4.4 *Statistical analysis and definition of study groups in the relevant randomised controlled trials*

4.4.1 Data imputation

In REALISTIC, PREDICT, DOSEFLEX, J-RAPID and HIKARI, missing data were imputed using the last observation carried forward (LOCF) or non-responder imputation (NRI) methods, for continuous and categorical outcome measures respectively. Of note, the NRI approach represents a very conservative imputation method for binary outcomes, compared with other approaches like the LOCF. The data imputation methods used in SWITCH were not reported by the study authors.[7]

4.4.2 REALISTIC

The REALISTIC trial randomised 1,063 participants 4:1 (851 CZP 200 mg Q2W vs. 212 PBO), with randomisation stratified by concomitant MTX use, prior TNFi exposure and disease duration to enable pre-specified analysis subgroups. The sample size for this study was calculated to have sufficient power not only for the primary outcome (in the overall population), but also in the subset of subjects with prior TNFi use history as stratified at randomisation.[73] A sample size of 419 subjects with previous TNFi use (335 CZP 200 mg Q2W vs. 84 PBO, with a 4:1 randomisation) was expected to achieve at least 90% power to show a statistically significant difference in the proportion of ACR20 responders at Week 12 between the CZP and PBO groups. This was true under the assumption that the percentage of

ACR20 responders in the PBO-treated group is [redacted] and the percentage of ACR20 responders in the CZP-treated group is at least [redacted] Both CZP and PBO were permitted to be administered as monotherapy, in combination with MTX with or without other cDMARDs, or with other cDMARDs excluding MTX.

Analyses at Week 12 were conducted in the overall full analysis set (FAS) and the FAS for each pre-specified subgroup. The FAS was defined as the intention-to-treat population (all randomised patients) during the double-blind phase. Withdrawals and missing data in the primary efficacy analysis (ACR20) were imputed using the NRI method. Treatment comparisons were performed using logistic regression with factors for treatment, concomitant use of MTX at baseline, previous TNFi use and disease duration (<2 vs. ≥2 years). Treatment effects were estimated with odds ratios (ORs) and 95% two-sided confidence intervals (CIs) obtained by fitting this model.[4]

For the analysis of secondary categorical end-points, treatment comparisons were performed using a similar logistic regression model as for the primary efficacy analysis. Treatment comparisons for change from baseline at Week 12 in ACR components were analysed using an analysis of covariance model (ANCOVA) with the same factors as for the primary efficacy model and baseline values as covariates. For categorical outcomes (ACR50/ACR70), missing data was imputed using the NRI approach. For continuous data (DAS28, HAQ-DI, fatigue, sleep index, and CDAI), missing data were imputed by the LOCF method, as were ACR component scores and EULAR response. LOCF-imputed ACR response data are provided in the appendix of this submission as well, for completeness. Comparisons were also performed for each pre-specified stratification subgroup (concomitant use of MTX at baseline, previous TNFi use and disease duration) for the primary endpoint only. Tests of interaction between treatment and each stratification variable were conducted separately at a P value of <0.05 significance level to examine whether treatment differences changed between each level of the assessed stratification variable. A significant interaction result implied that the treatment effect size (for the response variable) was influenced by the status of the assessed stratification variable. Of the outcomes presented in this submission, pre-specified statistical analyses were only performed in the TNFi experienced stratification group for the primary outcome measure (ACR response at Week 12). Statistical analyses of secondary outcomes within this stratification group (prior TNFi use) were conducted post-hoc, and therefore any p values presented for these endpoints are nominal and should be interpreted with caution.[4]

4.4.3 DOSEFLEX

The DOSEFLEX trial enrolled 333 patients into the open-label run-in and then randomised 209 participants 1:1:1 (69 CZP 200 mg Q2W + MTX vs. 70 CZP 400 mg Q4W + MTX vs. 69 PBO + MTX) at Week 18. With 67 subjects randomised per arm,

each pairwise comparison (CZP 400 mg Q4W vs. PBO and CZP 200 mg Q2W vs. PBO) would achieve at least 90% power to show a statistically significant difference in the percentage of ACR20 responders at Week 34 between the CZP dose groups and PBO. This was true under the assumption that the percent of ACR20 responders at Week 34 is 50% in the PBO group and 80% in the CZP-treated group.

Efficacy analyses presented for Week 16 outcomes (after the open-label run-in period) were conducted on the modified enrolled set (modified ES) of patients, defined as the number of patients who entered the run-in phase and received at least 1 dose of study drug. The FAS consisted of all treated, randomised subjects in the double-blind period. Efficacy analyses on the randomised population (Week 34) were conducted on the FAS.

Missing data for categorical variables (ACR responses and CDAI, SDAI, and DAS28(ESR) remission rates) were imputed using the NRI method, and for continuous variables (DAS28(ESR), fatigue, HAQ-DI, CDAI and SF-36 scores) using the LOCF method. LOCF-imputed ACR response data are provided in the appendix of this submission as well, for completeness. Subgroup analysis by prior TNFi use was post-hoc and therefore the resulting p value are nominal.[5]

4.4.4 PREDICT

The PREDICT study was designed to show that the RAPID3 is comparable to the CDAI in assessing response to CZP therapy after 12 weeks of treatment, as well as predicting long-term treatment success (Week 52 DAS28(ESR) remission) based on subjects' Week 12 responder status. Assuming that [redacted] of all subjects were Week 12 responders and [redacted] of Week 12 responders were also Week 52 LDA, a sample size of [redacted] subjects per group [redacted] would have approximately [redacted] Week 12 responders per group and the test would have approximately [redacted] power to reject the null hypothesis and demonstrate comparability of RAPID3. Thus [redacted] subjects would need to be enrolled in this study.

Efficacy analysis was conducted on the FAS population including all patients with a valid baseline and ≥1 post-baseline efficacy measurement. The precision of the effect estimates was improved by adjusting the differences in proportions for the primary endpoints for the baseline DAS28(ESR) scores, age, sex, disease duration (<2 or ≥2 years) and prior TNFi use. Missing data for categorical variables (eg. modified ACR responses) were imputed using the NRI method. However, LOCF-imputed ACR response data are provided in the appendix of this submission as well, for completeness. For continuous data, missing data were imputed by the LOCF approach.[6]

Of note, the mACR assessment used in the PREDICT study differed from the standard ACR assessment in two ways, as described in Appendix 8.6.2.

4.4.5 SWITCH

A sample size of 102 subjects with previous TNFi use (60 CZP vs. 30 PBO, after an assumed drop-out rate of 12%, with a 2:1 randomisation) was expected to achieve 80% power to show a statistically significant difference (at a 5% significance level) in the proportion of ACR20 responders at Week 12 between the CZP and PBO groups. However, after randomisation of 37 participants (27 CZP vs. 10 PBO) further enrolment was stopped as interim analyses indicated that the primary endpoint hypothesis had been confirmed and exposure of further patients to PBO was deemed unethical.[7]

4.4.6 Other relevant studies

4.4.6.1 J-RAPID

The sample size was based on an expected ACR20 response rate of 22% in the PBO group and at least 50% in the CZP 200 mg Q2W and 400 mg Q2W groups, as per previous clinical experience with CZP. Verification of superiority of the 200 mg and 400 mg doses over PBO for the primary end-point would then have 90% power at a 2-sided significance level of 2.5% in order to preserve the overall Type I error rate at α=0.05, with 71 patients per group (300 overall, pre-randomisation).

The primary end-point was assessed using the FAS of patients (who received at least 1 dose of study drug and provided efficacy data thereafter). ACR20 response comparisons between the CZP groups and PBO were carried out using logistic regression analysis with treatment group as a factor. Missing data for categorical variables (eg. ACR responses) were imputed using the NRI method. For continuous data, missing data were imputed by the LOCF approach.[8]

4.4.6.2 HIKARI

The sample size was based on an expected 20% ACR20 response in the PBO group and ≥42% in the CZP 200 mg Q2 W group, based on previous clinical experience in monotherapy trials. A projected 91 patients were needed in each group to detect the superiority of CZP 200 mg Q2W over PBO with 90% power at a two-sided significance level of 0.05. The target number of patients was set at 200 (100 patients per group) to allow for dropouts.

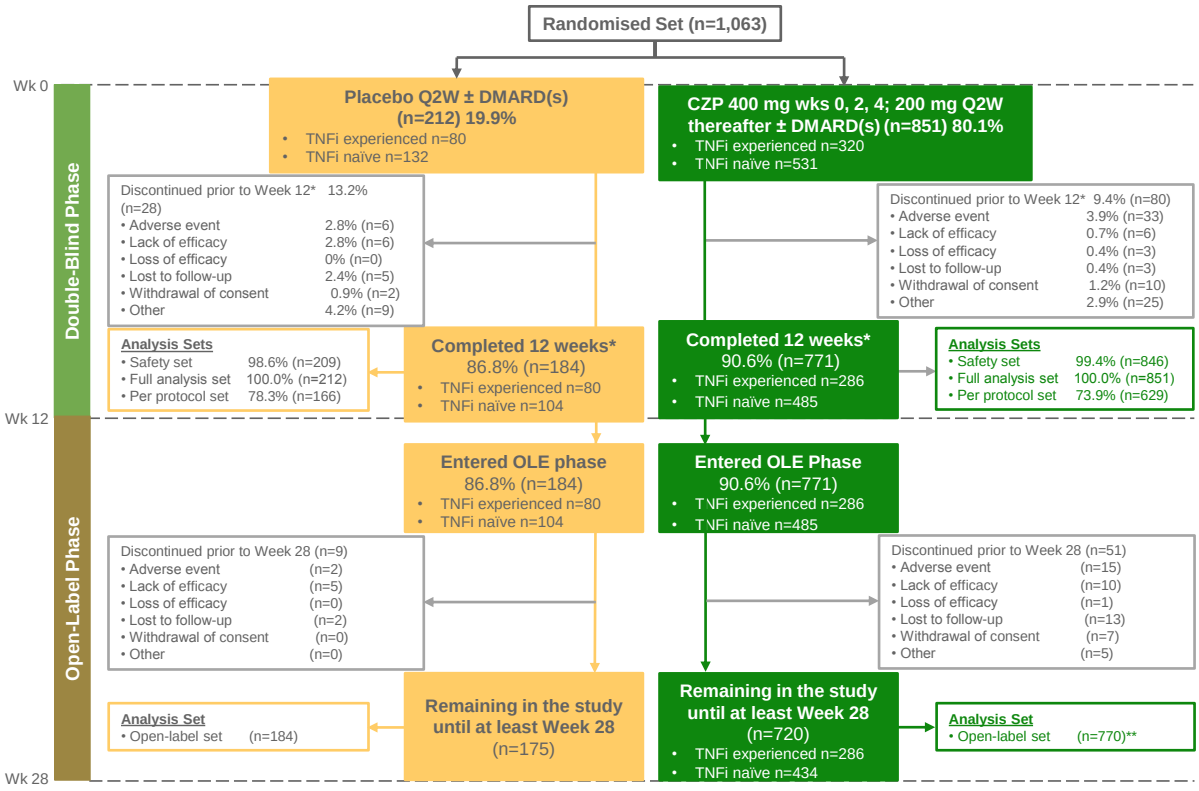
The primary end-point was assessed using the FAS of patients. ACR response comparisons between CZP 200 mg Q2W and PBO were carried out using logistic regression analysis with treatment group as factor. Missing data for categorical variables (eg. ACR responses) were imputed using the NRI method. For continuous data, missing data were imputed by the LOCF approach.[9]

4.5 *Participant flow in the relevant randomised controlled trials*

4.5.1 REALISTIC participant flow

A CONSORT diagram presenting the flow of participants enrolled into the full randomised set in the REALISTIC trial is presented in Figure 10. A total of 1,063 patients with active RA, according to the 1987 ACR criteria, were studied in REALISTIC. At Week 0, 851 patients were randomised to CZP 200 mg Q2W treatment and 212 patients were randomised to PBO. Of these patients, 400 had previous TNFi exposure; 320 were randomised to receive CZP 200 mg Q2W and 80 to receive PBO. Of the overall trial population, 771 (90.6%) CZP-treated patients and 184 (86.8%) PBO-treated patients completed double-blind treatment up to Week 12. All of these patients then entered the OLE. The most common reason for discontinuation of treatment was AEs.[4]

Figure 10: Patient disposition in REALISTIC



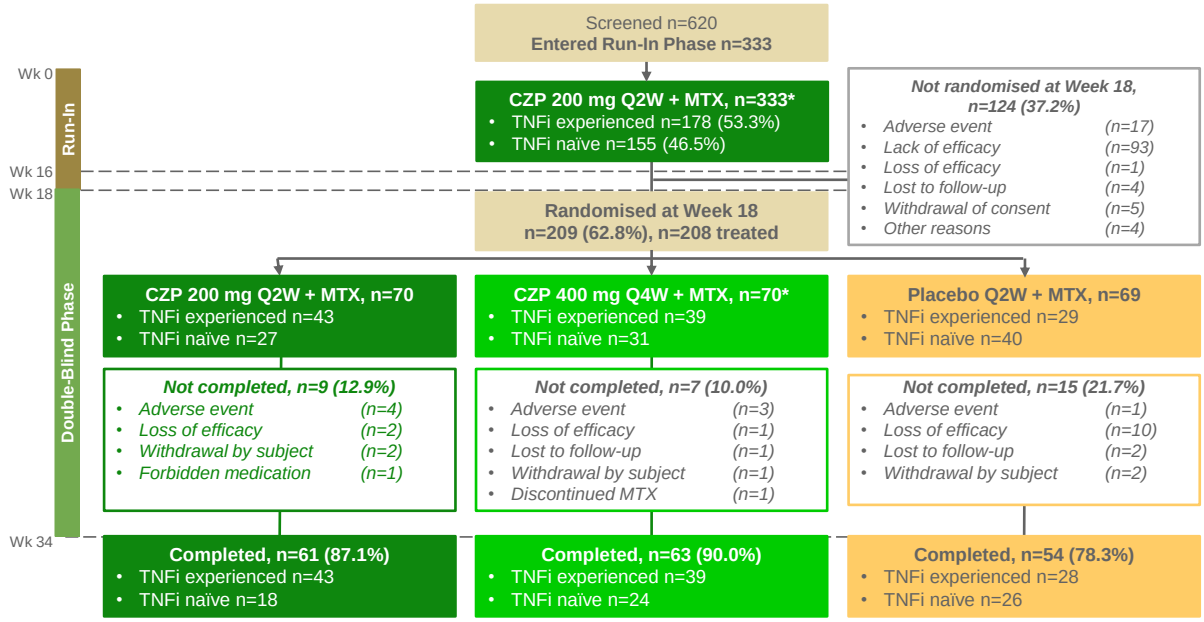
*Double-blind phase. **The number varies from n=771 CZP Week 12 completers entering the OLE due to 1 CZP completer who discontinued the OLE after Week 12 due to an AE, did not receive any study medication in the OLE, and was not included in the OLE analysis set

4.5.2 DOSEFLEX participant flow

A CONSORT diagram presenting the flow of participants in the DOSEFLEX trial is presented in Figure 11. A total of 333 patients were included in the 16 week open-

label run-in phase. Of these patients, 209 (62.8%) completed 16 weeks of CZP 200 mg Q2W treatment and were randomised for the 16 week double-blind phase commencing at Week 18. A total of 70 patients were randomised to continue CZP 200 mg Q2W during the double-blind phase, of which 43 (61.4%) had received prior TNFi therapy. A total of 70 patients were randomised to receive CZP 400 mg Q4W, of which 39 (55.7%) had received prior TNFi therapy and 69 patients were randomised to PBO Q2W, of which 29 (42.0%) had previously been treated with TNFi therapy. Of the overall trial population, 61 (87.1%) CZP 200 mg Q2W-treated patients, 63 (90.0%) CZP 400 mg Q4W-treated patients and 54 (78.3%) PBO-treated patients completed double-blind treatment up to Week 34. The most common reason for discontinuation of treatment was AEs.

Figure 11: Patient disposition in DOSEFLEX

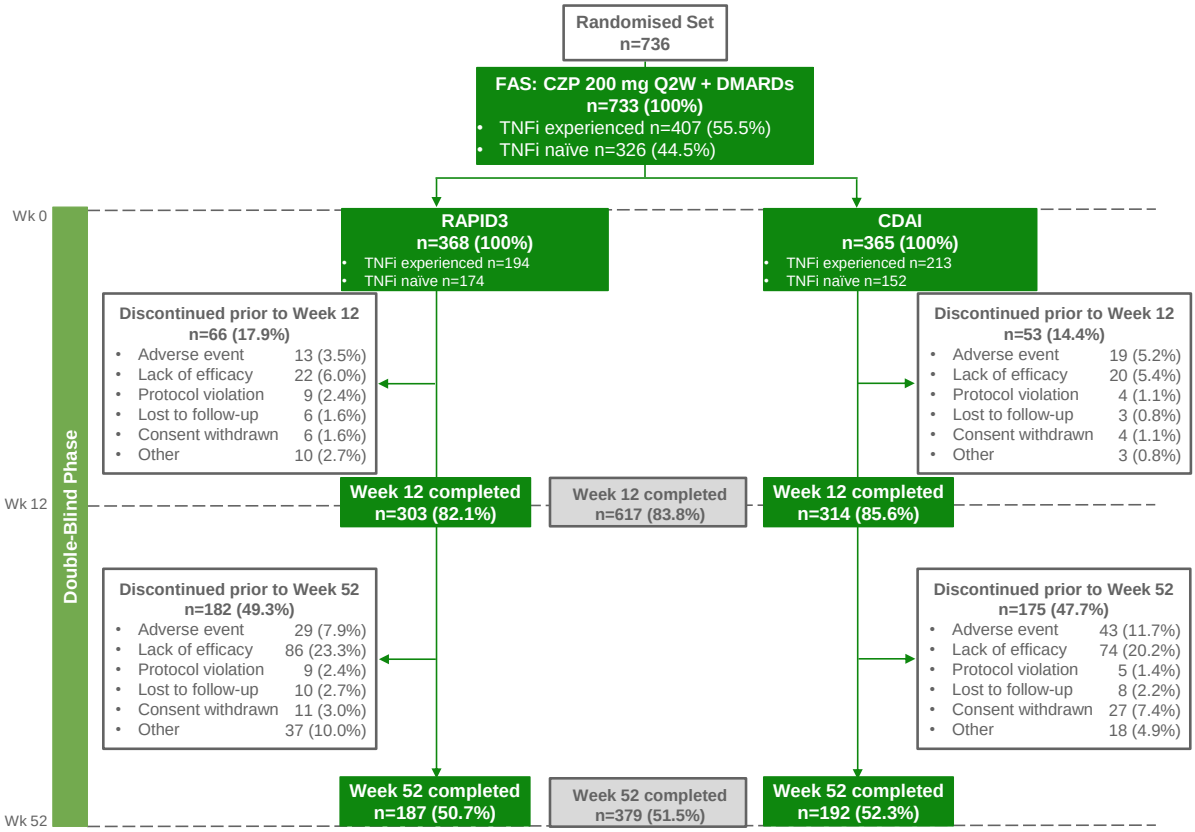


*One patient, randomised to the CZP 400 mg Q4W group, was not treated in the double-blind phase

4.5.3 PREDICT participant flow

A CONSORT diagram presenting the flow of participants in the PREDICT trial is presented in Figure 12. A total of 736 patients were randomised and 733 of these patients were included in the FAS, with 50.7% of the patients assigned to RAPID-3 and 52.3% of the patients assigned to CDAI completing the full 52 weeks of treatment. A total of 369 patients were randomised to RAPID-3 and 367 patients were randomised to CDAI. All 733 patients in the FAS received CZP 200 mg Q2W, of which 407 had received prior TNFi therapy. Of the overall trial population, 379 (51.5%) CZP 200 mg Q2W-treated patients completed double-blind treatment up to Week 52. The most common reason for discontinuation of treatment was AEs.

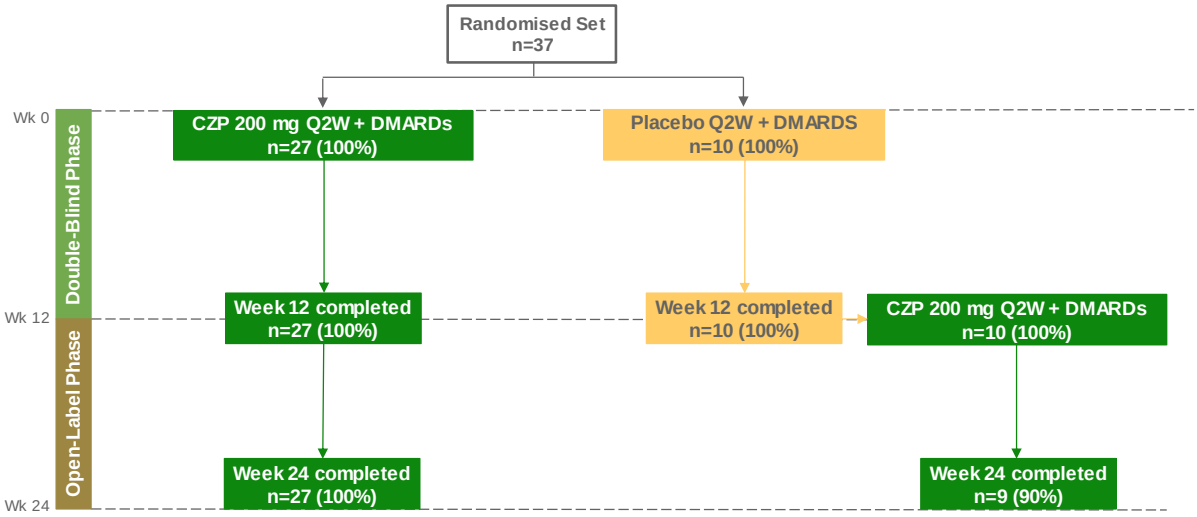
Figure 12: Patient disposition in PREDICT



4.5.4 SWITCH participant flow

A CONSORT diagram presenting the flow of participants in the SWITCH trial is presented in Figure 13. A total of 37 patients were randomised with 27 patients randomised to CZP 200 mg Q2W and 10 patients randomised to PBO. A total of 100% of the patients assigned to CZP and 100% of patients assigned to PBO completed the 12 week double-blind phase. At Week 12, all patients entered the open label phase to receive CZP 200 mg Q2W. All 27 patients initially randomised to CZP and 9 (90%) of the patients initially randomised to PBO completed the OLE.

Figure 13: Patient disposition in SWITCH

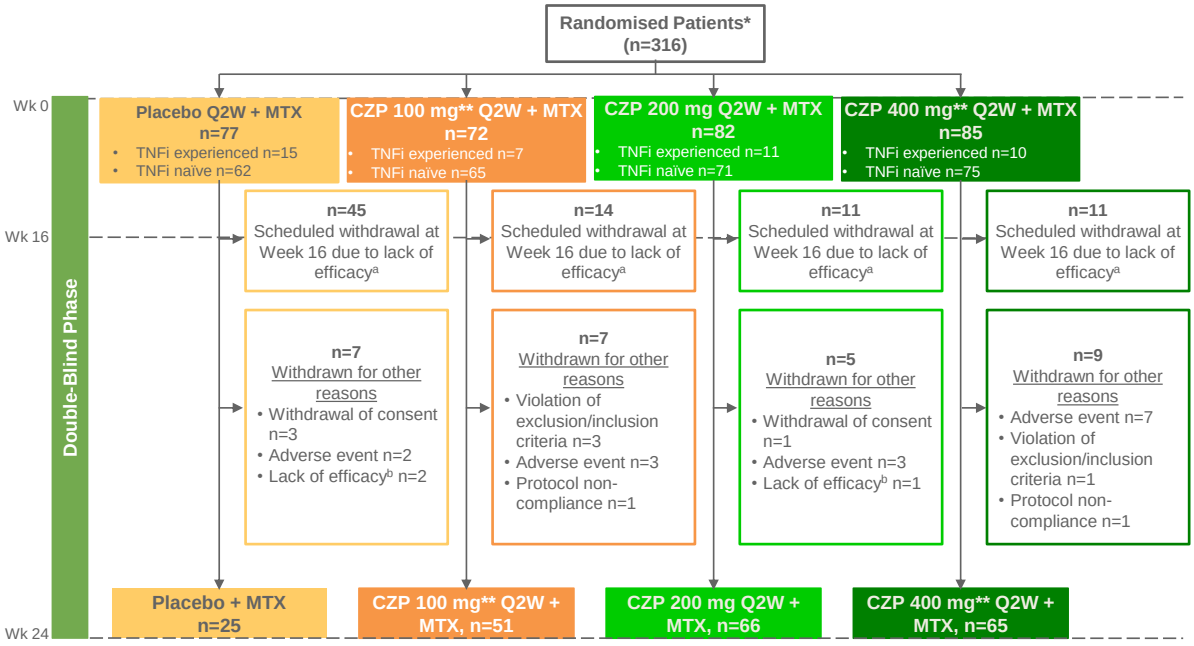


4.5.5 Other relevant studies

4.5.5.1 J-RAPID

A CONSORT diagram presenting the flow of participants in the J-RAPID trial is presented in Figure 14. A total of 316 patients were randomised, with 72, 82 and 85 patients randomised to CZP 100 mg Q2W, CZP 200 mg Q2W and CZP 400 mg Q2W, respectively, and 77 patients randomised to PBO. The scheduled withdrawal of patients not achieving an ACR20 response at Weeks 12 and 14 affected a total of 15% of the patients assigned to either of the CZP groups and 60% of patients assigned to PBO.[8]

Figure 14: Patient disposition in J-RAPID



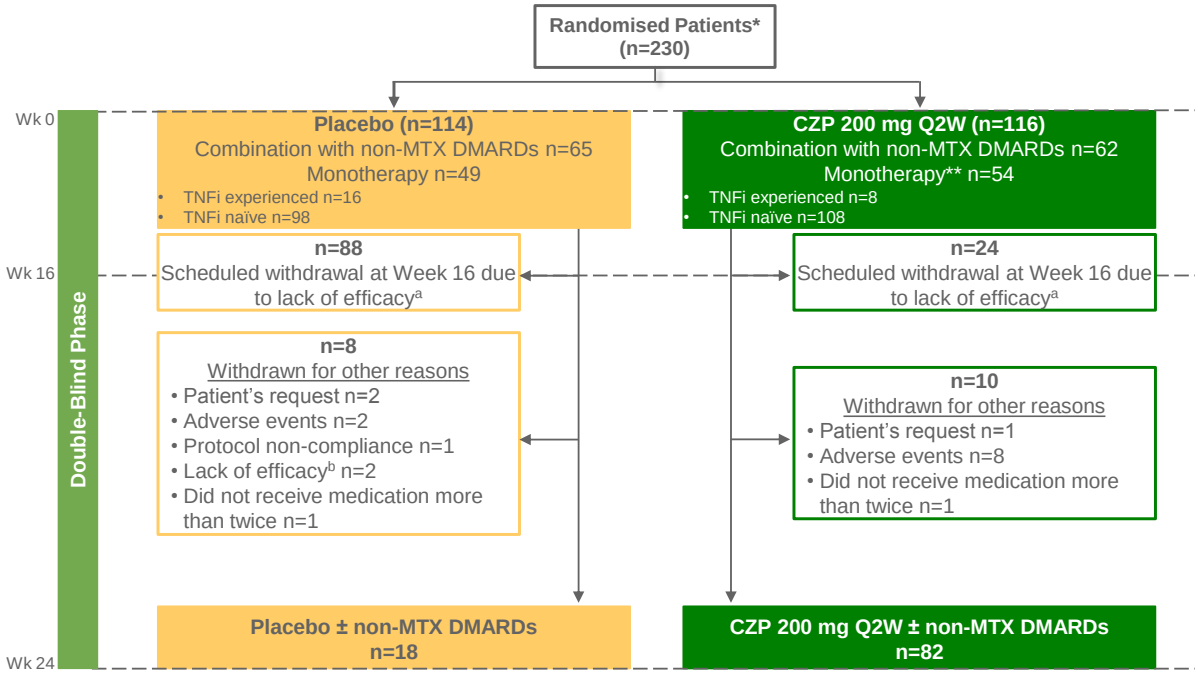
*Full analysis set

**For maintenance in RA, CZP 400 mg Q2W and 100 mg Q2W are not approved doses in the European Union.
aACR20 response was not achieved at Week 12 and Week 14.
bEfficacy of study drug was insufficient at times other than Week 12 and Week 14. Patients not showing an ACR20 response at Week 12 and Week 14 were withdrawn from the study at Week 16 and were eligible to enter an open-label extension, as were patients completing the study.
Figure adapted from Yamamoto 2014[8]

4.5.5.2 HIKARI

A CONSORT diagram presenting the flow of participants in the HIKARI trial is presented in Figure 15. A total of 230 patients were randomised with 116 patients randomised to CZP 200 mg Q2W and 114 patients randomised to PBO. The scheduled withdrawal of patients not achieving an ACR20 response at Weeks 12 and 14 affected a total of 21% of the patients assigned to CZP and 77% of patients assigned to PBO.[9]

Figure 15: Patient disposition in HIKARI



*Full analysis set
**CZP is indicated in combination with MTX in the European Union. CZP may be given as monotherapy in case of intolerance to MTX, or when continued treatment with MTX is inappropriate. CZP in combination with non-MTX cDMARDs is not approved in the European Union.
aACR20 response was not achieved at Week 12 and Week 14.
bEfficacy of study drug was insufficient at times other than Week 12 and Week 14. Patients not showing an ACR20 response at Week 12 and Week 14 were withdrawn from the study at Week 16 and were eligible to enter an open-label extension, as were patients completing the study.
Figure adapted from Yamamoto 20149

4.5.6 Baseline characteristics of participants

The baseline characteristics of the patients included the relevant RCTs that have been included in the analysis of the submission are compared in Table 13. Most patient baseline characteristics were comparable within the study groups of each trial.

- The REALISTIC trial included patients from a broader, clinically relevant population (including those with prior TNFi use, concomitant MTX treatment or treatment with a range of other concomitant DMARDs). Similar proportions of patients within the PBO and CZP treatment arms were using concomitant DMARDs at baseline, with 9.4% and 13.2% of PBO and CZP treated patients, respectively, using two or more concomitant DMARDs. A total of 69.2% and 67.5% of patients were using concomitant MTX at baseline, in the CZP and PBO arm, at mean weekly doses of 17.2 and 16.6 mg/week for CZP and PBO, respectively. The proportion of patients with TNFi experience was comparable between the PBO and CZP treatment arms.[4] Overall, the patients showed similar characteristics at baseline between the CZP and PBO treatment groups. At study baseline, patients had on average over 8 years disease duration (8.6 and 8.9 years, respectively for CZP and PBO groups) and high disease activity (mean DAS28(ESR) score of 6.4 in both groups).[4]
- The DOSEFLEX trial included a total of 333 patients for the 16-week run-in phase, of which about half (53.5%, n=178) had prior TNFi use. Patients with TNFi experience typically had a longer mean disease duration than TNFi naïve patients.[5] The patients randomised into PBO, 200 mg CZP or 400 mg CZP for the double-blind phase (Week 16) showed no clinically-relevant differences in baseline characteristics between groups.[5] Similarly, baseline characteristics between patients receiving the two different dosing arms of CZP who had prior TNFi experience were very comparable. In comparison to the PBO group, patients randomised to either CZP treatment arm had a higher proportion of TNFi experienced patients.[5]
- The PREDICT trial included 733 patients in the FAS, with 407 (55.5%) patients having previous TNFi experience. All patients were randomised to RAPID-3 or CDAI assessment arms with similar demographic and disease baseline characteristics between groups.[6] As expected, patients within these groups who had received prior TNFi therapy had a longer average disease duration than those in the overall study population.
- The SWITCH trial included 27 patients who had previously experienced secondary inadequate response or were intolerant to a TNFi other than CZP. These patients were randomised with CZP or PBO. The baseline characteristics were similar between these two treatment groups: on average patients had a high baseline disease activity and long-standing disease. Concomitant MTX use was reported at mean doses of 16.4 and 16.1 mg/week for the CZP and PBO groups, respectively.[7]
- The J-RAPID trial included Japanese patients with active RA that had previously shown inadequate response to treatment with MTX. All patients continued MTX treatment alongside CZP or PBO. There were no marked

differences in baseline characteristics between the groups.[8] A total of 13.4% of patients treated with CZP 200 mg Q2W had previous TNFi experience, as did 19.5% of PBO-treated patients.

- The HIKARI trial included Japanese patients with active RA which had shown inadequate response to at least one DMARD, including MTX. These patients were randomised for treatment with CZP or PBO in monotherapy or in combination with non-MTX DMARDs. Baseline characteristics were similar between CZP and PBO groups. Amongst CZP-treated patients, 6.9% had previously been treated with anti-TNF therapy, as has 14% of patients in the PBO arm.[9]

Table 13: Baseline characteristics of participants in the studies across treatment groups

		(n) [% TNFi experienced of total]	Mean age (SD), years	Female, n (%)	Mean disease duration, years (SD)	HAQ-DI mean (SD)	DAS28(ESR), mean (SD)	RF-positive (≥14 IU/mL), n (%)
REALISTIC[4] (NCT00717236)	All subjects	Overall patients (n=1,063)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		TNFi experienced (n=400) [37.6%]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	CZP 200 mg Q2W +/- MTX/cDMARDs	Overall patients (n=851)	55.4 (12.4)	660 (77.6)	8.6 (8.8)	1.5 (0.6)	6.4 (0.9)	555 (73.9)
		TNFi experienced (n=320) [37.6%]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	PBO +/- MTX/cDMARDs	Overall patients (n=212)	53.9 (12.7)	169 (79.7)	8.9 (9.1)	1.6 (0.6)	6.4 (0.9)	137 (76.5)
		TNFi experienced (n=80) [37.7%]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
DOSEFLEX[5] (NCT00580840)	All subjects‡	Overall patients (n=333)	54.2 (12.8)	76.0	6.4 (4.5)	1.52 (0.64)	6.4 (1.0)	315 (94.6)
		TNFi experienced (n=178) [53.5%]	54.2 (12.07)	77.0	7.6 (4.4)	1.6 (0.6)	6.4 (0.9)	167 (93.8)
	CZP 200 mg Q2W + MTX‡	Overall patients (n=70)	55.6 (10.7)	49 (70.0)	5.9 (4.2)	1.6 (0.7)	6.4 (0.8)	65 (92.9)
		TNFi experienced (n=43) [61.4%]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	CZP 400 mg Q4W + MTX‡	Overall patients (n=70)	53.1 (13.8)	58 (82.9)	6.4 (4.7)	1.4 (0.6)	6.2 (1.0)	65 (92.9)
		TNFi experienced (n=39) [55.7%]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	PBO + MTX‡	Overall patients (n=69)	51.5 (13.2)	56 (81.2)	6.5 (4.6)	1.4 (0.6)	6.4 (1.0))	69 (97.1)
		TNFi experienced (n=29) [42.0%]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

		(n) [% TNFi experienced of total]	Mean age (SD), years	Female, n (%)	Mean disease duration, years (SD)	HAQ-DI mean (SD)	DAS28(ESR), mean (SD)	RF-positive (≥14 IU/mL), n (%)
PREDICT[6] (NCT01255761)	CZP 200 mg Q2W +/- MTX/cDMARDs	Overall patients (n=733)	54.9	571 (77.9)	8.9 (9.1)	5.9†	6.3 (1.1)	493 (71.1)
		TNFi experienced (n=407) [55.5%]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		RAPID-3 assigned (n=368) [52.7%]	54.0	279 (75.8)	8.8 (9.3)	6.1†	6.3 (1.1)	251 (72.3)
		CDAI assigned (n=365) [58.4%]	55.7	292 (80.0)	9.1 (8.9)	5.8†	6.3 (1.1)	242 (69.9)
SWITCH[7] (NCT01147341)	CZP 200 mg Q2W + cDMARDs	TNFi experienced (n=27) [100.0%]	56.1	NR	12.0	1.5	5.5*	NR
	PBO + cDMARDs	TNFi experienced (n=10) [100.0%]	59.0	NR	14.0	1.1	5.4*	NR
J-RAPID[74] **(NCT00791999)**	CZP 200 mg Q2W + MTX	Overall patients (n=82[redacted]	50.6 (11.4)	69 (84.1)	5.6 (4.2)	1.1 (0.7)	6.2 (0.8)	11 (13.4)
	PBO + MTX	Overall patients (n=77) [redacted]	51.9 (11.1)	66 (85.7)	5.8 (4.1)	1.2 (0.7)	6.5 (0.9)	15 (19.5)
HIKARI[75] **(NCT00791921)**	CZP 200 mg Q2W -/+ non-MTX cDMARDs≠	Overall patients (n=116) [6.9%]	56.0 (10.2)	83.7 (71.6)	5.4 (4.0)	1.05 (0.7)	6.1 (0.9)	8 (6.9)
	PBO -/+ non-MTX cDMARDs	Overall patients (n=114) [14.0%]	55.4 (9.8)	88 (77.2)	5.8 (4.3)	1.21 (0.7)	6.3 (1.0)	16 (14.0)

SD: standard deviation; NR: not reported; CV: coefficient of variation; RAPID-3: routine assessment of patient index data 3; CDAI: clinical disease activity index. ŦFor REALISTIC, selected baseline characteristics were only recorded in a subset of the patients within the overall study population, and are indicated as [n] where appropriate. ‡For DOSEFLEX, baseline characteristics for “all subjects” refers to all subjects in the modified enrolled set who entered the 4 week run-in phase, while the PBO and CZP stratification data represent patients who completed the run-in phase and were subsequently randomised into the three treatment groups (PBO, 200 mg CZP or 400 mg CZP) for the double-blind phase. †For PREDICT, MD-HAQ Global scores at baseline (within a range of 0-10) are presented. *For SWITCH, DAS28(CRP) at baseline is presented, DAS28(ESR) was not measured. ≠CZP in combination with non-MTX cDMARDs is not approved in the European Union.

4.6 *Quality assessment of the relevant randomised controlled trials*

The full quality assessments of all included RCTs can be found in Appendix 8.4.2. A summary of the assessment can be found in Table 14.

Table 14: Quality assessment results for parallel group RCTs

Trial acronym	REALISTIC	J-RAPID	HIKARI	DOSEFLEX	PREDICT	SWITCH
Was randomisation carried out appropriately?	Yes	Yes	Yes	Yes	Yes	Not clear
Was the concealment of treatment allocation adequate?	Yes	Yes	Yes	Yes	Yes	Not clear
Were the groups similar at the outset of the study in terms of prognostic factors?	Yes	Yes	Yes	Yes	Yes	Yes
Were the care providers, participants and outcome assessors blind to treatment allocation?	Yes	Yes	Yes	Yes	Yes	Not clear
Were there any unexpected imbalances in drop-outs between groups?	No	No	No	No	No	No
Is there any evidence to suggest that the authors measured more outcomes than they reported?	No	No	No	No	No	No
Did the analysis include an intention-totreat- analy sis? If so, was this appropriate and were appropriate methods used to account for missing data?	Yes	Yes	Yes	Yes	Yes	Yes
	Adapted from Systematic reviews: CRD's guidance for undertaking reviews in health care (University of York Centre for Reviews and Dissemination)					

4.7 *Clinical effectiveness results of the relevant randomised controlled trials*

4.7.1 REALISTIC

This submission will present data for the population of patients with prior TNFi use from the REALISTIC study (n=400; CZP n=320 vs PBO n=80), in addition to the overall study population for ACR response. Additionally analyses were performed in subgroups of TNFi experienced patients, consisting of patients on monotherapy, or concomitant MTX therapy with or without other concomitant cDMARDs. A further patient subgroup, treated with CZP in combination with concomitant cDMARDs therapy excluding MTX, is not within the scope of this submission and thus data are not presented here. Where possible, kinetics are presented for the double-blind phase up to Week 12 and inset tables are used to present findings from the OLE up to Week 28. The OLE analyses presented here only consider patients randomised to CZP at Week 0.

For more information on trial design, please refer to Section 4.3.1.

4.7.1.1 Clinical effectiveness in the overall study population

4.7.1.1.1. Clinical responses (ACR20, ACR50, ACR70 rates)

In the overall study population, consisting of both TNFi naïve and experienced patients, the onset of CZP treatment effect was evident from as early as Week 2 (the first timepoint assessed), with the ACR20 response in the CZP group significantly higher than in the PBO group. This response was maintained until Week 12; the end of the double-blind treatment phase ($p<0.001$ at each time point) (Table 15). Similarly, a significantly greater proportion of CZP patients achieved an ACR50 response from Week 2 compared to PBO, which was maintained to Week 12 ($p<0.001$ at each time point) (Table 15). ACR70 response rates were significantly higher in the CZP group at Weeks 6 and 12 than in the PBO group ($p=0.02$ and $p<0.001$, respectively) (Table 15).

Efficacy analyses were also conducted on the OLE set, defined as patients who completed 12 weeks of double-blind CZP treatment and who received at least 1 dose of open-label CZP at Week 12 (“CZP 200 mg Q2W (OLE)”). For this population, the number of patients achieving an ACR20, ACR50 and ACR70 response increased between Week 12 and Week 28 (Table 15).

Table 15: REALISTIC study: ACR response rates in overall study population during 12 week double-blind phase and up to Week 28 in the OLE (NRI)

n (% response) p OR [95% CI][†]	PBO (n=212)[a]	CZP 200 mg Q2W (n=851)[a]	Wk 0 CZP 200 mg Q2W (OLE)[b] (n=770)
ACR20			
Week 2	[redacted]	[redacted]	[redacted]
Week 6	[redacted]	[redacted]	[redacted]
Week 12	[redacted] (25.9%)	[redacted] (51.1%) p<0.001	[redacted]
Week 28	[redacted]	[redacted]	[redacted]
ACR50			
Week 2	[redacted]	[redacted]	[redacted]
Week 6	[redacted]	[redacted]	[redacted]
Week 12	[redacted] (9.9%)	[redacted] (26.6%) p<0.001	[redacted]
Week 28	[redacted]	[redacted]	[redacted]
ACR70			
Week 2	[redacted]	[redacted]	[redacted]
Week 6	[redacted]	[redacted]	[redacted]
Week 12	[redacted] (2.8%)	[redacted] (12.9%) p<0.001	[redacted]
Week 28	[redacted]	[redacted]	[redacted]

[a]FAS, NRI. [b]Open Label Set, NRI. The OLE group only includes patients randomised to CZP at Week 0, who completed 12 weeks of double-blind phase before entering the OLE.
OR: odds ratio; CI: confidence interval. P-values indicated where calculated.
[†]In FAS, CZP 200 mg Q2W versus PBO

The overall population of patients from the REALISTIC trial contains those patients who received CZP or PBO monotherapy and those who received treatment in combination with MTX (with or without other cDMARDs). The ACR responses of patients within these subgroups are presented in Appendix 8.7.1.1. Overall, a similar ACR response rate was seen with CZP in these subgroups in the overall trial population.

4.7.1.2 Clinical effectiveness in the TNFi experienced population

Summary of clinical effectiveness in the TNFi experienced population of REALISTIC

ACR response rate improvements are comparable between the overall study population and TNFi experienced patients.

- ACR20 and ACR50 response rates were significantly higher in CZP versus PBO treatment groups at Week 12, both for the overall study population and the TNFi experienced subgroup.
- The effects seen at Week 12 were maintained up to Week 28, in CZP monotherapy as well as in combination with MTX.

Disease activity, fatigue and sleep problems as well as physical function were numerically improved with CZP treatment versus PBO, in monotherapy and in combination with MTX.

- At Week 12, numerically greater reductions were seen in DAS28(ESR) and DAS28(CRP) scores, CDAI score, fatigue and sleep symptom scores and HAQ-DI scores with CZP treatment versus PBO, representing clinically meaningful improvements.
- These improvements were maintained until Week 28 and similar in both CZP monotherapy and in combination with MTX.
- DAS28(ESR) remission rates followed the same trend as DAS28(ESR) score reductions, however, the proportion of patients in remission was numerically greater in patients treated with MTX, compared to monotherapy, within both the PBO and the CZP treatment arms.

4.7.1.2.1. Clinical responses

ACR response rates

In the population of patients with prior TNFi experience in the REALISTIC trial, the ACR20 response rates at Week 12 were statistically significantly higher for the CZP group compared with the PBO group (47.2% vs 27.5%, respectively, $p<0.01$), as shown in Figure 16. The differences between the CZP and PBO groups were also significant for the ACR50 response rates (21.5% vs 11.3%, respectively, $p<0.05$) (Figure 17). There was a numerically greater proportion of ACR70 responders at Week 12 in the CZP group compared to PBO-treated patients (9.1% vs 3.8%, respectively), however this did not reach a statistically significant difference (Figure 18).[4]

Efficacy analyses were also conducted on the OLE set, defined as patients who completed 12 weeks of double-blind CZP treatment and who received at least 1 dose of open-label CZP at Week 12. For these patients, the initial effect seen at Week 12 was further maintained up to Week 28 (Figure 16, Figure 17 and Figure 18).

The effect of CZP on ACR improvement seen in the TNFi experienced population is comparable to the effect seen in the overall study population (Figure 16, Figure 17, Figure 18 and Table 15). In both the overall study population and the subset of patients with prior TNFi experience, both ACR20 and ACR50 response rates at Week 12 are significantly higher in the CZP groups than the PBO groups, indicating a similar efficacy of CZP in patients with or without prior TNFi exposure.[4]

ACR response rates presented in this section are reported as per the NRI imputation method. LOCF-imputed data are available in Appendix 8.7.2. The kinetics of ACR response rates were comparable for the NRI and LOCF datasets.

Figure 16: REALISTIC study: kinetics of ACR20 response rates in TNFi experienced population during 12 week double-blind phase (OLE rates in inset table) (NRI)

Figure 17: REALISTIC study: kinetics of ACR50 response rates in TNFi experienced population during 12 week double-blind phase (OLE rates in inset table) (NRI)

Figure 18: REALISTIC study: kinetics of ACR70 response rates in TNFi experienced population during 12 week double-blind phase (OLE rates in inset table) (NRI)

The overall population of TNFi experienced patients from the REALISTIC trial includes those patients who received CZP or PBO monotherapy, and those who received treatment in combination with MTX (with or without other cDMARDs). The ACR responses of patients within these subgroups are presented in Appendix 8.7.1.2. Overall, a similar ACR response rate was seen with CZP in these subgroups in the TNFi experienced population. Accordingly, at Week 12, CZP treatment in combination with MTX achieved significantly higher ACR20 and ACR50 response rates than PBO in this TNFi experienced patient subgroup. Notably, at Week 12, a high [redacted] ACR50 response rate was recorded for the PBO monotherapy group, comparable with CZP monotherapy and MTX combination subgroups.

ACR component scores

The numerically greater proportion of patients with an ACR response at Week 12 in the CZP group than the PBO is reflected in the ACR component scores at baseline, Week 12 and Week 28 in the OLE. For each ACR component, there was a numerically greater reduction in component score between baseline and Week 12 in the CZP group compared to PBO. The reduction in all component scores in the CZP group was maintained at Week 28 in the OLE (see Appendix 8.7.2.4).

The numerically greater reduction in component scores in the CZP compared with the PBO group in the TNFi experienced population is also reflected in the monotherapy and combination with MTX subgroups (Appendix 8.7.2.5).

EULAR response

The proportion of EULAR good and moderate responders in the CZP group at Week 12 was numerically greater than that in the PBO group. At Week 12, [redacted] of patients in the PBO group had a good EULAR response whereas [redacted] of patients treated with CZP in the FAS and OLE sets, respectively, had good EULAR responses. Within the OLE, by Week 28 the proportion of good responders had increased compared to Week 12, with corresponding reductions in the proportion of moderate and non-responders (Table 16).

Table 16: REALISTIC study: EULAR response rates in the TNFi experienced population (LOCF)

		PBO[a] (n=80)	CZP 200 mg Q2W[a] (n=320)	Wk 0 CZP 200 mg Q2W (OLE; n=286)[b]
EULAR Response, n (%)				
Week 12	Good	[redacted]	[redacted]	[redacted]
	Moderate	[redacted]	[redacted]	[redacted]
	None	[redacted]	[redacted]	[redacted]
Week 28	Good	[redacted]	[redacted]	[redacted]
	Moderate	[redacted]	[redacted]	[redacted]
	None	[redacted]	[redacted]	[redacted]

[a]FAS (LOCF); [b]Open label set (LOCF), The OLE group only includes patients randomised to CZP at Week 0, who completed 12 weeks of double-blind phase before entering the OLE. No p-values were calculated.

The proportion of EULAR good and moderate responders in the TNFi experienced population was numerically similar in the monotherapy and combination with MTX TNFi experienced patients (Appendix 8.7.1.3).

4.7.1.2.2. Disease activity

DAS28 scores

Reductions in the disease activity, as measured by the DAS28(ESR) and DAS28(CRP) scores, were numerically greater in the CZP than the PBO group from Week 2 of the double-blind phase and maintained to Week 12 (mean DAS28(CRP) change from baseline of -1.64 for patients treated with CZP 200 mg Q2W versus -1.02 for those treated with PBO, whilst the DAS28(ESR) change from baseline was -1.79 for patients treated with CZP 200 mg Q2W versus -1.13 for those treated with PBO [no statistical tests were performed]).[4]

The initial reductions in the disease activity with CZP were further continued or maintained until Week 28 (Figure 19 and Figure 20).

Figure 19: REALISTIC study: kinetics of DAS28(ESR) score in TNFi experienced population during 12 week double-blind phase (OLE scores in inset table) (LOCF)

Figure 20: REALISTIC study: kinetics of DAS28(CRP) score in TNFi experienced population during 12 week double-blind phase (OLE scores in inset table) (LOCF)

The DAS28(ESR) and DAS28(CRP) scores of TNFi experienced patients on a monotherapy regimen or in combination with MTX (with or without other cDMARDs) are presented in Appendix 8.7.1.4. The reductions in both DAS28(ESR) and DAS28(CRP) were numerically greater in the CZP-treated patients compared to PBO-treated patients in both subgroups. Overall, the initial effect seen at Week 12 was further maintained or continued up to Week 28 in the group randomised to CZP at Week 0, for both subgroups.

DAS28(ESR) remission

At Week 12, a numerically greater proportion of CZP patients were in remission compared to patients in the PBO group. In the OLE, the proportion of patients in remission further increased between Weeks 12 and 28 (Table 17).

Table 17: REALISTIC study: DAS28(ESR) remission in TNFi experienced population during 12 week double-blind phase and OLE (NRI)

		PBO[a] (n=80)	CZP 200 mg Q2W[a] (n=320)	Wk 0 CZP 200 mg Q2W (OLE; n=286)[b]
DAS28(ESR) Remission, n (%)				
Week 12	Remitter			
	Non-remitter			
Week 28	Remitter			
	Non-remitter			

[a]FAS (NRI); [b]Open label set (NRI), The OLE group only includes patients randomised to CZP at Week 0, who completed 12 weeks of double-blind phase before entering the OLE. No p-values were calculated.

The increase in the proportion of CZP patients in remission at Week 12 in comparison to PBO, as well as at Week 28, was reflected in the subgroups of patients treated with monotherapy or therapy in combination with MTX (with or without other cDMARDs). The proportion of CZP patients in remission within the monotherapy and combination with MTX subgroups was greater at Week 12 than the proportion of responders in the PBO subgroups, and further increased by Week 28. However, the proportion of patients in remission (both CZP- and PBO-treated) was greater in the combination with MTX subgroup than the monotherapy subgroup (Appendix 8.7.1.5).

CDAI scores

The disease activity as measured through the CDAI score in the CZP group reduced by a greater amount between Weeks 0 and 12 than in the PBO group. This initial reduction in CDAI score in the CZP treated patients continued to Week 28 in the OLE (Figure 21).

Figure 21: REALISTIC study: CDAI in TNFi experienced population during 12 week double-blind phase (OLE scores in inset table) (LOCF)

The CDAI scores across all treatment groups at Week 0 were comparable for patients treated as monotherapy, or in combination with MTX (with or without other cDMARDs). Similar, improvements in CDAI scores between Weeks 0 and 12 during the double-blind phase and up to Week 28 in the OLE were seen within these subgroups (Appendix 8.7.1.6).

4.7.1.2.3. Impact on fatigue and sleep

Fatigue

Fatigue was assessed by Fatigue Assessment Scale. Subjects reported their level of fatigue (tiredness) by answering the following question: “Please rate your fatigue (weariness, tiredness) during the past 7 days, on a scale of 0 to 10” where 0 is ‘No Fatigue’ and 10 is ‘Fatigue as bad as you can imagine’.

CZP treatment was associated with a clinically meaningful reduction (≥1 point improvement) in fatigue as early as Week 2, and maintained to Week 12, with fatigue scores numerically lower than the PBO group at all time points. This clinically

meaningful reduction in fatigue was maintained with CZP treatment in the OLE to Week 28 (Figure 22).

Figure 22: REALISTIC study: Patient's Assessment of Fatigue in TNFi experienced population during 12 week double-blind phase (OLE scores in inset table) (LOCF)

Similar clinically meaningful reductions from baseline in fatigue were seen with CZP in the monotherapy and combination with MTX (with or without other cDMARDs) subgroups at Week 12, and were further maintained to Week 28 in the OLE (Appendix 8.7.1.7).

Sleep

CZP treatment was associated with a clinically meaningful reduction (≥6 point improvement) in sleep symptoms from baseline between Weeks 0 and 6. This reduction was maintained at Week 12 and was numerically greater than PBO (Figure 23). These clinically significant reductions in sleep problems with CZP continued in the OLE until Week 28.

Figure 23: REALISTIC study: Sleep Problem Index II score in TNFi experienced population during 12 week double-blind phase (OLE scores in inset table) (LOCF)

Similar clinically meaningful reductions from baseline in sleep symptoms were seen with CZP in monotherapy and in combination with MTX (with or without other cDMARDs) subgroups at Week 12, and further maintained to Week 28 in the OLE (Appendix 8.7.1.8).

4.7.1.2.4. Impact on physical function (HAQ-DI)

CZP treatment resulted in numerical reductions of HAQ-DI score as early as Week 2, which were maintained to Week 12 in the double-blind phase (Figure 24), and also through to Week 28 in the OLE. Of note, the mean HAQ-DI score of the CZP patients was numerically lower than PBO at baseline. The reductions of HAQ-DI score were numerically greater for patients undergoing treatment with CZP compared to PBO-treated patients throughout the double-blind phase of the study and by Week 12. The treatment interaction between HAQ-DI score in patients without TNFi experience and those with TNFi experience was found to be significant (interaction $p<0.05$), indicating improved physical function in patients with no prior TNFi experience.[4]

Figure 24: REALISTIC study: HAQ-DI to Week 12 for TNFi experienced population during 12 week double-blind phase (OLE scores in inset table) (LOCF)

HAQ-DI scores were comparable between the entire TNFi experienced population, and the monotherapy, and combination with MTX (with or without cDMARD) subgroups (Appendix 8.7.1.9).

4.7.2 DOSEFLEX

During the open-label run-in phase, all patients received 400 mg CZP at Weeks 0, 2 and 4, followed by CZP 200 mg Q2W up to Week 16 in combination with MTX. Patients who had achieved an ACR20 response at Week 16 were randomised at Week 18 1:1:1 to receive 200 mg CZP Q2W, 400 mg CZP Q4W, or PBO up to Week 34, in combination with MTX.[5]

This submission will present data for the population of patients with prior TNFi use only except for ACR response, which will, as a reference, also be presented for the overall study population. The TNFi experienced population includes one patient who received monotherapy pooled with data from patients who received treatment in combination with MTX.

Data from Week 12 is presented separately in Appendix 8.8.

For more information on trial design, please refer to Section 4.3.2.

4.7.2.1 Clinical effectiveness in the overall study population

4.7.2.1.1. Clinical responses (ACR20, ACR50, ACR70 rates)

In the overall study population of DOSEFLEX, consisting of both TNFi naïve and experienced patients, the proportion of ACR20 responders at Week 34 was significantly higher in both the CZP 200 mg Q2W group and the CZP 400 mg Q4W group compared to the PBO group (p=0.009 and p=0.017, respectively) (Table 18). The ACR50 response rate was also significantly higher at Week 34 in both CZP groups than the PBO group (p<0.05 for comparisons against both dosing arms). Additionally, the ACR70 response rate in the CZP 400 mg Q4W group was significantly greater than in the PBO group (p=0.005). Although the ACR70 response rate in the CZP 200 mg Q2W group was numerically greater than that of the PBO group, there was no statistically significant difference (p=0.052).[5]

Table 18: DOSEFLEX study: ACR response rates in overall study population during the first 34 weeks of study (NRI)

	PBO + MTX (n=69)	CZP 200 mg Q2W + MTX (n=70)	CZP 400 mg Q4W + MTX (n=69)
ACR20: n (% response rate)			
Open label run-in phase[a]			
Week 4	[redacted]	[redacted]	[redacted]
Week 16	[redacted]	[redacted]	[redacted]
Double-blind phase[b]			
Week 24	[redacted]	[redacted]	[redacted]
Week 34	(44.9%)	(67.1%)**	(65.2%)*
ACR50: n (% response rate)			
Open label run-in phase[a]			
Week 4	[redacted]	[redacted]	[redacted]
Week 16	[redacted]	[redacted]	[redacted]
Double-blind phase[b]			
Week 24	[redacted]	[redacted]	[redacted]
Week 34	(30.4%)	(50.0%)*	(52.2%)*
ACR70: n (% response rate)			
Open label run-in phase[a]			
Week 4	[redacted]	[redacted]	[redacted]
Week 16	[redacted]	[redacted]	[redacted]
Double-blind phase[b]			
Week 24	[redacted]	[redacted]	[redacted]
Week 34	(15.9%)	(30.0%)	(37.7%)*

*p<0.05, **p<0.01. [a]Open-label run-in phase: all groups, including PBO, received CZP in the 16-week open-label run-in phase, [b]Double-blind phase
FAS, NRI.

4.7.2.2 Clinical effectiveness in the TNFi experienced population

Summary of clinical effectiveness in the TNFi experienced population of DOSEFLEX

At Week 34, the end of the double-blind phase, ACR response rates were numerically greater in both CZP treatment groups than the CZP to PBO group and comparable to those seen in the overall study population.

- Maintenance of response was similar, regardless of CZP dosing schedule, for most efficacy parameters.
- ACR component scores followed the same trend, showing maintenance of numerically greater improvements in symptoms, in CZP versus CZP to PBO treatment arms within the double-blind phase.
- EULAR response rates revealed a majority of patients with moderate and good response in the 200 mg Q2W and 400 mg Q4W CZP treatment arms, respectively, compared to a majority of non-responders following randomisation to PBO.

Initial improvements in DAS28(ESR) scores, CDAI scores, HAQ-DI scores and SF-36 component scores at Week 16 were maintained in CZP-treated patients to Week 34 but worsened in the PBO group.

- The two different CZP dosage regimes achieved comparable improvements at Week 34, reflected in numerically greater DAS28(ESR) remission rates in CZP versus PBO treated patients.
- Patients' assessment of fatigue scores were numerically lower in CZP groups than PBO group at Week 16. The two CZP groups maintained improvements until Week 34 while a worsening was noticed in the PBO group.

4.7.2.2.1. Clinical responses

ACR response rates

In the population of patients with prior TNFi experience in the DOSEFLEX trial, the kinetics of ACR response rates in response to CZP treatment from Baseline to Week 16, the end of the CZP run-in phase, were comparable in all three groups. ACR20 responders to CZP treatment were then randomised to treatment with PBO, CZP 200 mg Q2W or 400 mg Q4W, resulting in maintained high ACR response rates with CZP. In the subsequent double-blind phase, Weeks 16 to 34, the ACR response rates in both CZP treatment arms were maintained to numerically greater than those in the PBO group. At Week 34, ACR20 response was significantly higher in CZP 200 mg Q2W versus PBO (nominal p=0.003), and numerically improved in the CZP 400 mg Q4W versus PBO group (nominal p=0.057) (see Figure 25, Figure 26, Figure 27).

The CZP effect on ACR response rate kinetics in the TNFi experienced population was comparable to the one seen in the overall study population (Table 18). In both the overall study population and the subset of patients with prior TNFi experience, the ACR20/50/70 are numerically similar at all time points.

ACR response rates presented in this section are reported using NRI imputation. LOCF-imputed data are available in Appendix 8.7.2. The kinetics of ACR response rates were comparable between imputation methods, with numerically higher response rates reported for LOCF versus NRI data, especially for ACR20 response rates.

Figure 25: DOSEFLEX study: kinetics of ACR20 response rates in TNFi experienced population during the first 34 weeks of study (NRI)

Figure 26: DOSEFLEX study: kinetics of ACR50 response rates in TNFi experienced population during the first 34 weeks of study (NRI)

Figure 27: DOSEFLEX study: kinetics of ACR70 response rates in TNFi experienced population during the first 34 weeks of study (NRI)

ACR component scores

The ACR component scores in both CZP groups and PBO reduced in the run-in phase between Weeks 0 and 16. After Week 16, those patients assigned to CZP treatment showed numerically greater reductions across all ACR component scores than the PBO group, indicating potentially greater improvements in symptoms of CZP-treated patients than those treated with PBO during the double-blind phase (Appendix 8.8.3.1).

EULAR response rate

The proportion of EULAR good and moderate responders in the CZP treatment groups and the PBO treatment group was similar across all three groups at Week 16, but by Week 34, the majority of patients in the CZP treatment groups were at least moderate responders, with [redacted] of CZP 200 mg Q2W and [redacted] of 400 mg Q4W patients classed as either good or moderate responders, respectively. In contrast, in the PBO group at Week 34, [redacted] of patients were classed as good and moderate responders, respectively (Table 19).

Table 19: DOSEFLEX study: EULAR response rates in the TNFi experienced population (LOCF)

		CZP 200 mg Q2W + MTX* (n=42)	CZP 400 mg Q4W + MTX* (n=39)	PBO + MTX* (n=27)
EULAR Response, n (%)				
Week 16*	Good	[redacted]	[redacted]	[redacted]
	Moderate	[redacted]	[redacted]	[redacted]
	None	[redacted]	[redacted]	[redacted]
Week 34	Good	[redacted]	[redacted]	[redacted]
	Moderate	[redacted]	[redacted]	[redacted]
	None	[redacted]	[redacted]	[redacted]

*All patients received CZP 200 mg Q2W during the run-in phase from Week 0–16
FAS (LOCF). No p-values were calculated.

4.7.2.2.2. Disease activity

DAS28 scores

During the 16-week open-label run-in phase, DAS28(ESR) scores improved across all patient groups. The mean DAS28(ESR) scores in all three treatment groups at Week 16 were comparable. By Week 34, the initial reductions (ie improvements) in disease activity scores at Week 16 had been maintained in the two CZP treatment groups, whereas a numerical worsening in disease activity was seen in patients randomised to PBO. The scores were comparable between CZP groups at Week 34 (Figure 28).

Figure 28: DOSEFLEX study: kinetics of DAS28(ESR) score in TNFi experienced population during the first 34 weeks of study (LOCF)

DAS28(ESR) remission

Similar to the trend seen in DAS28(ESR) scores, patients in all groups (initially treated with CZP 200 mg Q2W) had comparable proportions in remission following the run-in phase at Week 16. By Week 34, the proportion of patients in remission was numerically greater in both CZP treatment arms in comparison to those randomised to PBO (Table 20).

Figure 29: DOSEFLEX study: DAS28(ESR) remission in TNFi experienced population (LOCF)

		CZP 200 mg Q2W + MTX* (n=43)	CZP 400 mg Q4W + MTX* (n=39)	PBO + MTX* (n=28)
DAS28(ESR) Remission, n (%)				
Week 16*	Remitter	[redacted]	[redacted]	[redacted]
	Non-remitter	[redacted]	[redacted]	[redacted]
Week 34	Remitter	[redacted]	[redacted]	[redacted]
	Non-remitter	[redacted]	[redacted]	[redacted]

*All patients received CZP 200 mg Q2W during the run-in phase from Week 0–16
FAS (LOCF). No p-values were calculated.

CDAI scores

The disease activity, as measured through CDAI across all treatment groups, reduced during the CZP run-in phase between Weeks 0 and 16, at which point the scores in all three groups were comparable. Following randomisation to Week 34, these initial reductions were maintained in both CZP groups, however the disease activity score worsened in those randomised to the PBO group (Figure 30).

Figure 30: DOSEFLEX study: kinetics of CDAI score in TNFi experienced population during the first 34 weeks of study (LOCF)

4.7.2.2.3. Impact on fatigue and sleep

Patient-reported fatigue scores, as assessed with the 10-point Fatigue Assessment Scale, improved during the CZP run-in phase between Week 0 and Week 16 across patient groups. The scores were similar in all groups at the end of the open-label run-in phase (Week 16). The scores were maintained and comparable between CZP groups at Week 34, however, worsened in those randomised to PBO (Figure 31).

Figure 31: DOSEFLEX study: kinetics of Patient's Assessment of Fatigue scores in TNFi experienced population during the first 34 weeks of study (LOCF)

4.7.2.2.4. Impact on physical function (HAQ-DI)

Improvements in physical function were seen in all patient groups in the 16-week open-label run-in phase, with similar mean HAQ-DI scores reached by all groups at Week 16. By Week 34, these initial improvements in physical function were maintained in both CZP groups, however HAQ-DI score increased (ie worsened) in those randomised to PBO between Week 16 and 34. The scores were comparable between CZP groups at Week 34 (Figure 32).

Figure 32: DOSEFLEX study: kinetics of HAQ-DI score for TNFi experienced population during the first 34 weeks of study (LOCF)

4.7.2.2.5. Impact on patient health-related quality of life (SF-36)

Initial improvements from Baseline to Week 16 were seen in both the mental (MCS) and physical (PCS) component scores of the SF-36 following CZP treatment in the run-in phase. These improvements were maintained to Week 34 in both CZP groups, whereas they worsened in those randomised PBO (Figure 33, Figure 34).

Figure 33: DOSEFLEX study: SF-36 physical component scores (PCS) for TNFi experienced population at Weeks 0, 16 and 34 of study (LOCF)

Figure 34: DOSEFLEX study: SF-36 mental component scores (MCS) for TNFi experienced population at Weeks 0, 16 and 34 of study (LOCF)

The improvements from Week 0 to Week 16 seen in the overall PCS and MCS scores were also seen in the 8 domains of the SF-36. The improvement between Weeks 16 and 34 in both CZP treatment groups was maintained, however a worsening was seen in those randomised to the PBO group between Weeks 16 and 34 (Table 20).

Table 20: DOSEFLEX study: SF-36 domain scores for the TNFi experienced population at Weeks 0, 16 and 34 of study (LOCF)

SF-36 domain scores [n] mean (SD)	PBO + MTX* (n=29)	CZP 200 mg Q2W + MTX* (n=43)	CZP 400 mg Q4W + MTX* (n=39)
Physical Functioning			
Week 0	[redacted]	[redacted]	[redacted]
Week 16	[redacted]	[redacted]	[redacted]
Week 34	[redacted]	[redacted]	[redacted]
Role Physical			
Week 0	[redacted]	[redacted]	[redacted]
Week 16	[redacted]	[redacted]	[redacted]
Week 34	[redacted]	[redacted]	[redacted]
Bodily Pain			
Week 0	[redacted]	[redacted]	[redacted]
Week 16	[redacted]	[redacted]	[redacted]
Week 34	[redacted]	[redacted]	[redacted]
General Health			
Week 0	[redacted]	[redacted]	[redacted]
Week 16	[redacted]	[redacted]	[redacted]
Week 34	[redacted]	[redacted]	[redacted]
Vitality, mean			
Week 0	[redacted]	[redacted]	[redacted]
Week 16	[redacted]	[redacted]	[redacted]
Week 34	[redacted]	[redacted]	[redacted]
Social Functioning			
Week 0	[redacted]	[redacted]	[redacted]
Week 16	[redacted]	[redacted]	[redacted]
Week 34	[redacted]	[redacted]	[redacted]
Role Emotional			
Week 0	[redacted]	[redacted]	[redacted]
Week 16	[redacted]	[redacted]	[redacted]
Week 34	[redacted]	[redacted]	[redacted]
Mental Health			
Week 0	[redacted]	[redacted]	[redacted]
Week 16	[redacted]	[redacted]	[redacted]
Week 34	[redacted]	[redacted]	[redacted]

*Final group sizes were PBO (n=28), CZP 200 mg Q2W (n=43), CZP 400 mg Q4W (n=39) at the end of Week 34. All patients received CZP 200 mg Q2W during the run-in phase from Week 0–16

FAS (LOCF). No p-values were calculated. Numbers in square brackets indicate number of patients included in the analysis at each timepoint.

4.7.3 PREDICT

This submission will present data for the population of patients with prior TNFi use only (n=407), except for mACR response[2] which will also be presented for the overall study population. Additionally, analyses were performed in subgroups of TNFi experienced patients, consisting of patients undergoing monotherapy, concomitant MTX therapy (with or without other cDMARDs), or concomitant cDMARD therapy excluding MTX. Analyses of the latter subgroup are not within the scope of this submission and are not presented here.

For more information on trial design, please refer to Section 4.3.3.

4.7.3.1 Clinical effectiveness in the overall study population

4.7.3.1.1. Clinical responses (mACR20, mACR50, mACR70 rates)

In the overall study population of PREDICT, consisting of both TNFi naïve and experienced patients, the mACR response rates at Week 12 were [redacted] mACR20 responders, [redacted] mACR50 responders and [redacted] mACR70 responders. The mACR20 and mACR50 response rates gradually decreased until Week 52. The mACR70 response was maintained between Weeks 12 and 52, ending at 19.1% at Week 52 (Table 21).

[2]The modified ACR assessment differs from the standard assessment by using a 28 joint count for disease activity measurements as well as MD-HAQ for patient-assessed physical function.

Table 21: PREDICT study: mACR20/50/70 response rates in overall study population during 52 week double-blind phase (NRI)

	CZP 200 mg Q2W (n=733)
mACR20: n (% response rate)	
Week 2	[redacted]
Week 12	[redacted]
Week 24	[redacted]
Week 52	[redacted]
mACR50: n (% response rate)	
Week 2	[redacted]
Week 12	[redacted]
Week 24	[redacted]
Week 52	[redacted]
mACR70: n (% response rate)	
Week 2	[redacted]
Week 12	[redacted]
Week 24	[redacted]
Week 52	[redacted]

FAS, NRI.

4.7.3.2 Clinical effectiveness in the TNFi experienced population

Summary of clinical effectiveness in the TNFi experienced population of PREDICT

The mACR response rates of the TNFi experienced population were similar to those in the overall study population, with peak response rates at Week 12 and typically numerically similar proportions of responders at all timepoints.

- Comparable mACR response rates were achieved by patients treated with CZP monotherapy and CZP in combination with MTX.
- All components of the mACR scores showed a similar trend to the mACR response rates.
- At all timepoints, mACR component scores were similar in the MTX combination group and the monotherapy group.

At Week 12, moderate EULAR response rates were reported in the majority of patients within the CZP treatment arm.

- EULAR responses continued to be achieved in the longer term up to Week 52 and responses were similar between monotherapy and combination with MTX subgroups.

Rapid improvements in disease activity and physical function were seen in the CZP monotherapy as well as MTX combination group, which were further maintained to Week 52.

- DAS28(ESR) and CDAI scores as well as MD-HAQ scores were reduced as early as Week 2 with further numerical reductions until Week 12.
- DAS28(ESR) remission rates followed a similar trend, also in patients treated with CZP monotherapy and in combination with MTX.

Workplace and household productivity showed rapid improvements which were maintained to Week 52 and comparable in both CZP subgroups with or without concomitant MTX use.

4.7.3.2.1. Clinical responses

mACR response rates

In the population of patients with prior TNFi experience in the PREDICT trial, the mACR response rates at Week 12 were [redacted] mACR20 responders, [redacted] mACR50

responders and [redacted] mACR70 responders. The mACR20 and mACR50 response rates gradually decreased until Week 52, at which timepoint the mACR20 response rate was [redacted] and mACR50 rate was [redacted] The mACR70 response rate was maintained between Weeks 12 and 52, ending at 16.0% at Week 52 (Figure 35). These proportions are numerically similar to those of the overall study population (containing both TNFi experienced and naïve patients) at all time points (Table 21).

mACR response rates presented in this section are reported as per the NRI imputation method. LOCF-imputed data are available in Appendix 8.9.2. The kinetics of mACR response rates were comparable between imputation methods, with numerically higher response rates reported for LOCF versus NRI data, especially for mACR20 response.

Figure 35: PREDICT study: kinetics of mACR20/50/70 response rates in TNFi experienced population during 52 week double-blind phase (NRI)

The mACR responses of patients within the subgroups of CZP monotherapy and combination with MTX are presented in Appendix 8.9.1.1. Overall, mACR responses in these subgroups were comparable to each other, and to responses of the TNFi experienced population.

EULAR response rates

At Week 12, the majority of patients had a moderate EULAR response. By Week 24, the proportion of moderate responders had slightly decreased, with a corresponding

increase in EULAR good responders (from [redacted] at Week 12 and 24, respectively). At Week 52, the proportion of good and moderate responders remained stable (Table 22).

Table 22: PREDICT study: EULAR response rates in the TNFi experienced population (LOCF)

		CZP 200 mg Q2W (n=398)
EULAR Response, n (%)		
Week 12	Good	[redacted]
	Moderate	
	None	
Week 24	Good	
	Moderate	
	None	
Week 52	Good	
	Moderate	
	None	

FAS (LOCF)

A similar trend in EULAR response was observed in the TNFi experienced CZP patients treated with monotherapy or in combination with MTX (Appendix 8.9.1.2).

4.7.3.2.2. Disease activity

DAS28 scores

Reductions in disease activity were seen following CZP treatment as early as Week 2 which were further continued to Week 12, (mean DAS28(ESR) score numerically reduced from baseline to Week 12. These initial improvements remained stable between Weeks 12 and 52 (Figure 36).

Figure 36: PREDICT study: kinetics of DAS28(ESR) score in TNFi experienced population during 52 week double-blind phase (LOCF)

The DAS28(ESR) scores of TNFi experienced patients on a monotherapy regimen, or treated in combination with MTX, are presented in Appendix 8.9.1.3. Overall, the reductions in disease activity in these subgroups were comparable to each other and to those of the TNFi experienced population up to Week 52.

DAS28(ESR) remission

The proportion of patients with DAS28(ESR) remission slightly increased between Weeks 12 and 52 (Table 23).

Table 23: PREDICT study: DAS28(ESR) remission in TNFi experienced population during 12 week double-blind phase and OLE (NRI)

		CZP 200 mg Q2W (n=407)
DAS28(ESR) Remission, n (%)		
Week 12	Remitter	
	Non-remitter	
Week 24	Remitter	
	Non-remitter	
Week 52	Remitter	
	Non-remitter	

FAS (NRI)

In the subgroups of TNFi experienced patients treated with monotherapy and in combination with MTX, the pattern of a gradual increase in the proportion of patients in remission over time was evident (Appendix 8.9.1.4).

CDAI scores

A fast reduction in disease activity as measured through the CDAI score in the overall TNFi experienced CZP treated population was seen as early as Week 2 and this continued to Week 12. These initial reductions were maintained between Weeks 12 and 52 (Figure 37).

Figure 37: PREDICT study: kinetics of CDAI score in TNFi experienced population during 52 week double-blind phase (LOCF)

The CDAI score in the CZP monotherapy and CZP in combination with MTX groups were comparable to each other and the TNFi experienced population (Appendix 8.9.1.5).

4.7.3.2.3. Impact on physical function (MD-HAQ)

Large improvements in physical function as assessed by MD-HAQ, were seen as early as Week 2 and continued to Week 12; these improvements were further maintained between Weeks 12 and 52 (Figure 38).

Figure 38: PREDICT study: kinetics of MD-HAQ global score for TNFi experienced population during 52 week double-blind phase (LOCF)

Improvements in MD-HAQ scores were comparable between the TNFi experienced population, and the TNFi experienced CZP monotherapy and CZP in combination with MTX subgroups (Appendix 8.9.1.6).

4.7.3.2.4. Impact on workplace and household productivity (WPS-RA)

Amongst the 38.8% of patients with prior TNFi treatment who were employed outside the home at study baseline, a high burden of disease was observed on both workplace and household productivity.[77] Improvements in workplace productivity amongst TNFi experienced patients were reported in the form of reductions in absenteeism (number of paid work days missed per month due to arthritis; Table 24) and presenteeism (paid work days with productivity reduced by ≥50% due to arthritis; Table 24). Similar improvements were seen in the level of arthritis interference with work productivity (Table 24). These initial improvements were maintained to Week 52.

Table 24: PREDICT study: Workplace productivity (WPS-RA) in the TNFi experienced population (Employed patients only, LOCF)

	CZP 200 mg Q2W (n=407)
Workplace productivity (employed patients only)	
Paid work days missed per month, [N] mean (SD)	
Week 0	[redacted]
Week 12	[redacted]
Week 24	[redacted]
Week 52	[redacted]
Paid work days with productivity reduced ≥50% per month, [N] mean (SD)[a]	
Week 0	[redacted]
Week 12	[redacted]
Week 24	[redacted]
Week 52	[redacted]
Arthritis interference with paid work productivity (0–10 scale), [N] mean (SD)[b]	
Week 0	[redacted]
Week 12	[redacted]
Week 24	[redacted]
Week 52	[redacted]

FAS (LOCF)
[a]Days with productivity reduced ≥50% per month do not include days counted in the previous question (full days missed).
[b]For 0-10 point scales: 0=no interference, 10=complete interference.

Similar improvements were noticed in terms of household productivity and participation in social activities. Large reductions were seen in the number of household work days missed per month due to arthritis, in household work days with productivity reduced by ≥50% due to arthritis and in the number of family, social and leisure activity days missed per month (Table 25). Similar reductions were seen in the level of arthritis interference on household productivity (Table 25). These initial improvements were maintained to Week 52.

Table 25: PREDICT study: Household productivity and participation in social activities (WPS-RA) in the TNFi experienced population (LOCF)

	CZP 200 mg Q2W (n=407)
Household productivity and participation in social activities	
Household work days missed per month, [N] mean (SD)	
Week 0	[redacted]
Week 12	[redacted]
Week 24	[redacted]
Week 52	[redacted]
Household work days with productivity reduced ≥50% per month, [N] mean (SD)[a]	
Week 0	[redacted]
Week 12	[redacted]
Week 24	[redacted]
Week 52	[redacted]
Family, social and leisure activity days missed per month, [N] mean (SD)	
Week 0	[redacted]
Week 12	[redacted]
Week 24	[redacted]
Week 52	[redacted]
Arthritis interference with household work productivity (0–10 scale), [N] mean (SD)[b]	
Week 0	[redacted]
Week 12	[redacted]
Week 24	[redacted]
Week 52	[redacted]

FAS (LOCF)
[a]Days with productivity reduced ≥50% per month do not include days counted in the previous question (full days missed).
[b]For 0-10 point scales: 0=no interference, 10=complete interference.

4.7.4 SWITCH

4.7.4.1 Clinical effectiveness in the TNFi experienced population

Summary of clinical effectiveness in the TNFi experienced population of SWITCH

CZP treatment of the TNFi experienced SWITCH trial population resulted in achievement of a significant ACR20 response at Week 12 compared to PBO.[7]

- ACR50 and ACR70 were numerically higher at Week 12 in CZP versus PBO as well.

- Patients switched from PBO to CZP in the OLE achieved substantial improvements in ACR responses, to match those initially randomised to CZP.

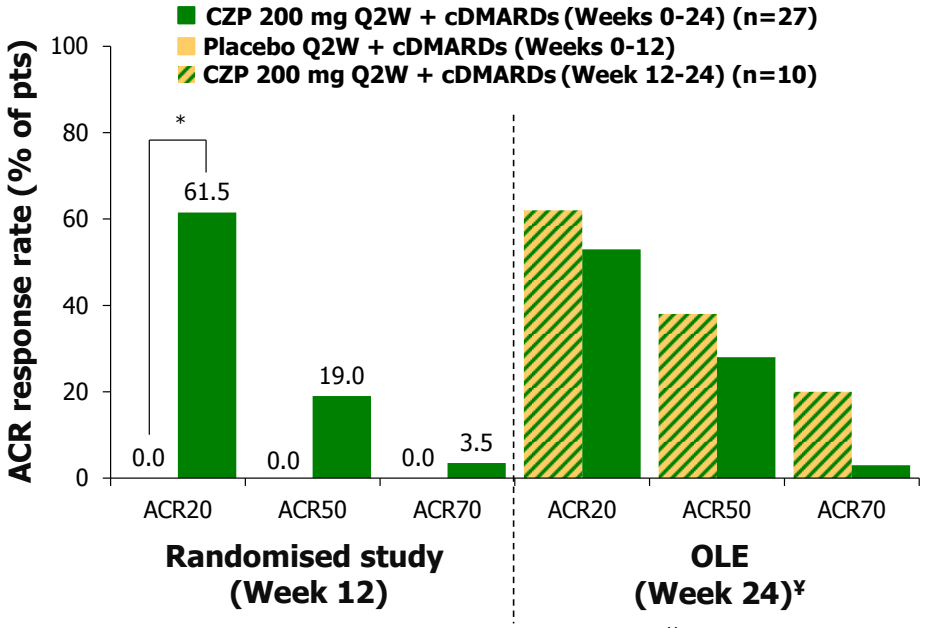
Disease activity and physical function significantly improved with CZP in the double-blind phase at Week 12, compared to PBO, as well as in the patient group switched from PBO to CZP in the OLE at Week 24.[7]

- DAS28(CRP) scores were significantly improved in CZP versus PBO treated patients at Week 12 (p<0.0005).
- The percentage of patients with improved HAQ-DI score was significantly higher (p=0.046) in CZP versus PBO treated patients at Week 12, and PBO patients switched to CZP during the OLE experienced HAQ-DI improvements as well.

4.7.4.1.1. Clinical responses

In the SWITCH trial, 61.5% of CZP patients achieved an ACR20 response at Week 12, whereas no PBO patients achieved ACR20 (p<0.005). In addition, the ACR50 and ACR70 responses were markedly higher at Week 12 in the CZP group. During the OLE phase, those patients who switched from PBO to CZP treatment demonstrated significant improvement in ACR20, ACR50 and ACR70. The patients who were initially randomised to CZP reached peak effectiveness by Week 12 (Figure 39).[7]

Figure 39: SWITCH study: ACR20/50/70 response rates in TNFi experienced (overall) population at Week 12 (randomised study) and Week 24 (OLE)

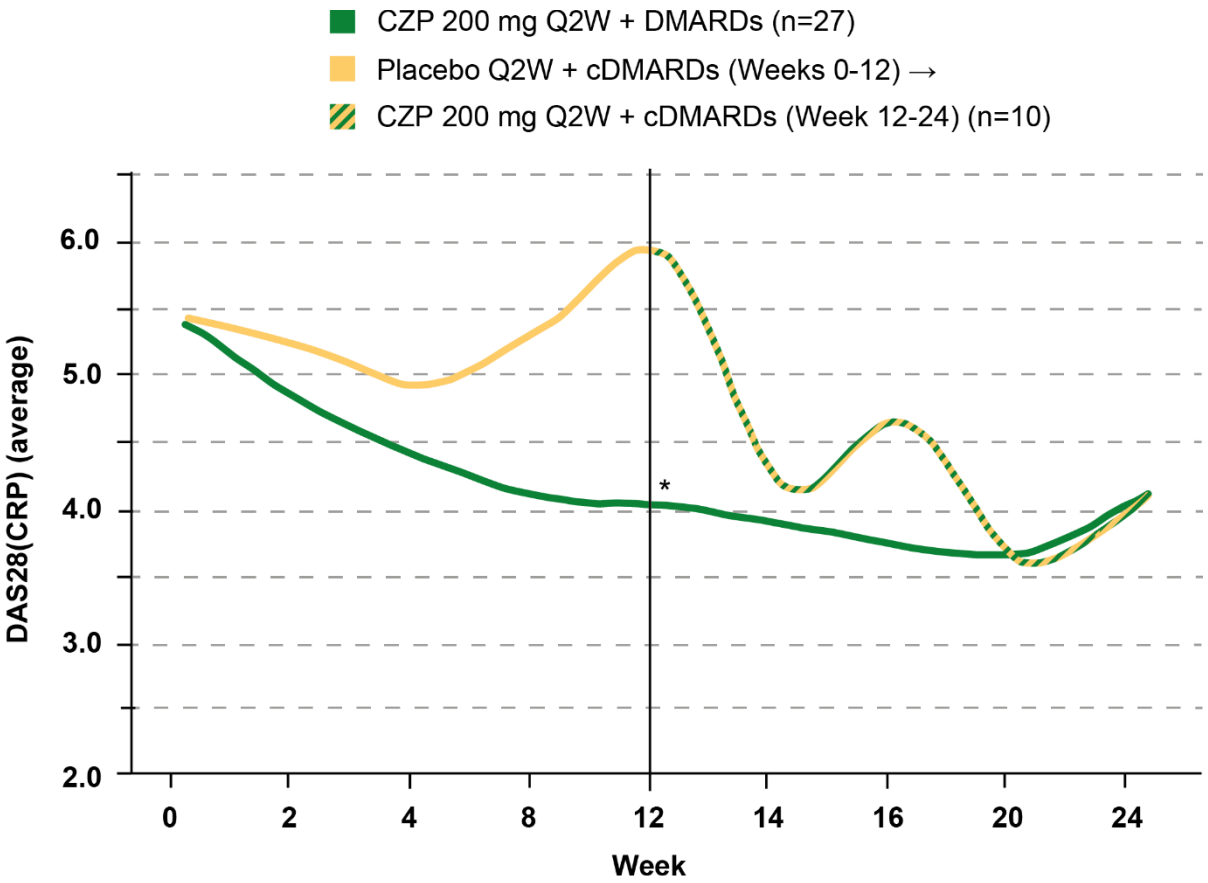


*p<0.005 CZP vs PBO. ACR rates for PBO were 0 at Week 12. ¥Data for Week 24 have been re-drawn from the manuscript. Data point values are not available.
Figure adapted from Schiff et al. 2014[7]

4.7.4.1.2. Disease activity

Significant reductions in disease activity were seen in the CZP treated group, with DAS28(CRP) scores significantly (p<0.0005) lower in CZP treated patients after 12 weeks as compared to patients initially treated with PBO. During the 12 weeks of the OLE phase (ie. from Week 12 to 24), the group of patients who switched from PBO to CZP treatment demonstrated increased improvement in DAS28 scores (Figure 40).[7]

Figure 40: SWITCH study: kinetics of DAS28(CRP) score in TNFi experienced (overall) population during 24 weeks of study

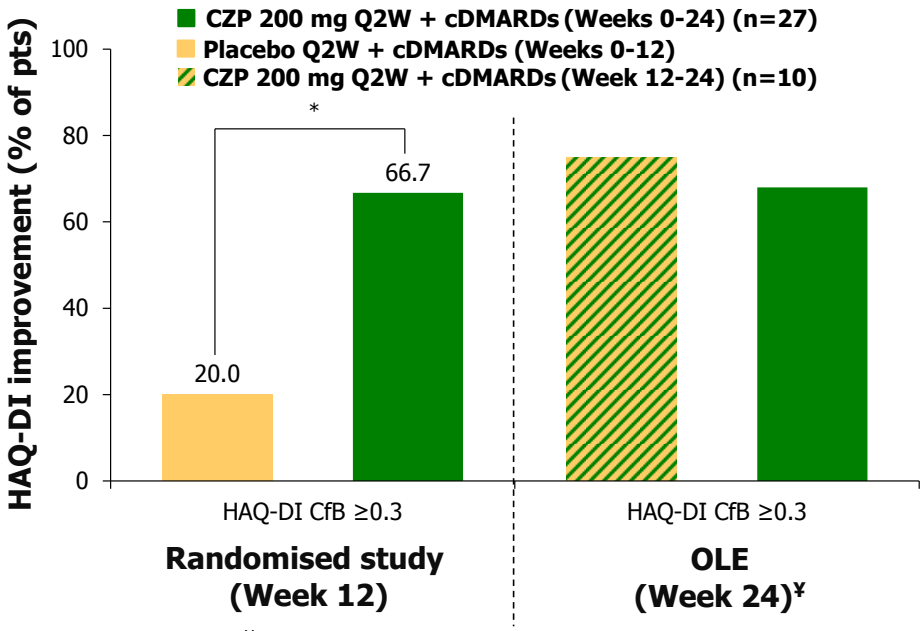


*p<0.0005, CZP vs PBO at Week 12.
Figure adapted from Schiff et al. 2014[7]

4.7.4.1.3. Impact on physical function (HAQ-DI)

The percentage of patients with physical function (HAQ-DI) improvement (defined as a decrease in HAQ-DI of ≥0.3) was 66.7% of the CZP patients versus 20% of the PBO-treated patients (p=0.046). During the OLE phase, the PBO patients who switched to CZP showed a greater HAQ-DI improvement by Week 24 (Figure 41).[7]

Figure 41: SWITCH study: HAQ-DI improvement for TNFi experienced (overall) population at Weeks 12 and 24 of OLE phase



*p=0.046 CZP vs PBO. ¥Data for Week 24 have been re-drawn from the manuscript. Data point values are not available.
Figure adapted from Schiff et al. 2014[7]

4.7.5 Other relevant studies

4.7.5.1 J-RAPID

Summary of clinical effectiveness in the TNFi experienced population of J-RAPID

While numbers were small in this subgroup of patients in the J-RAPID study, numerical improvements in clinical responses were observed in TNFi experienced patients treated with CZP.

- Numerical improvements of ACR response rates were reported with CZP 200 mg Q2W versus PBO, comparable to responses previously reported for the overall study population.
- Similarly, moderate to good EULAR response rates were reported for CZP 200 mg Q2W, with increases in responders from Weeks 12 to 24.

At Weeks 12 and 24, remission of disease activity was only achieved by patients within the CZP treatment arms, with no patients in the PBO group achieving remission.

4.7.5.1.1. Clinical responses

ACR response

Overall, the ACR responses of the TNFi experienced population in J-RAPID were numerically improved for patients taking CZP in comparison to those taking PBO (Table 26). Response rates in the TNFi experienced population were typically numerically

similar to those reported for the overall study population, reported elsewhere.[8] Of note, given the small patient sizes in the TNFi experienced population, any comparisons should be made with caution.

Table 26: J-RAPID study: ACR response rates of the TNFi experienced population during the 24-week double-blind phase (NRI)

	PBO + MTX (n=15)	CZP 200 mg Q2W + MTX (n=11)
ACR20: n (% response rate)		
Week 12	[redacted]	[redacted]
Week 24	[redacted]	[redacted]
ACR50: n (% response rate)		
Week 12	[redacted]	[redacted]
Week 24	[redacted]	[redacted]
ACR70: n (% response rate)		
Week 12	[redacted]	[redacted]
Week 24	[redacted]	[redacted]

FAS (NRI). No p-values calculated.

EULAR response

Within the CZP 200 mg Q2W treatment groups, the proportion of good and moderate responders increased between Weeks 12 and 24 whereas this proportion decreased in the PBO group. Of note, the proportion of good responders at both Weeks 12 and 24 numerically increased with CZP dose (Table 27).

Table 27: J-RAPID study: EULAR response rates of the TNFi experienced population during the 24 week double-blind phase (LOCF)

		PBO + MTX (n=15)	CZP 200 mg Q2W + MTX (n=11)
EULAR Response, n (%)			
Week 12	Good	[redacted]	[redacted]
	Moderate	[redacted]	[redacted]
	None	[redacted]	[redacted]
Week 24	Good	[redacted]	[redacted]
	Moderate	[redacted]	[redacted]
	None	[redacted]	[redacted]

FAS (LOCF). No p-values calculated.

4.7.5.1.2. Disease activity

DAS28(ESR) remission

At Week 12, there were no patients in remission in the CZP 200 mg Q2W treatment group. By Week 24[redacted] patient achieving remission with CZP 200 mg Q2W, while [redacted] patients in the PBO group entered remission (Table 28).

Table 28: J-RAPID study: DAS28(ESR) remission of the TNFi experienced population during the 24 week double-blind phase (LOCF)

		PBO + MTX (n=15)	CZP 200 mg Q2W + MTX (n=11)
DAS28(ESR) remission, n (%)			
Week 12	Remission	[redacted]	[redacted]
	Not remission	[redacted]	[redacted]
Week 24	Remission	[redacted]	[redacted]
	Not remission	[redacted]	[redacted]

FAS (LOCF). No p-values calculated.

4.7.5.2 HIKARI

All data presented below for HIKARI refer to the subgroups of TNFi experienced patients receiving PBO or CZP in monotherapy only.

Summary of clinical effectiveness in the TNFi experienced population of HIKARI receiving monotherapy

Within the small TNFi experienced patient subpopulation, numerical improvements in clinical response were observed in CZP versus PBO treatment.

- ACR response rates for TNFi experienced patients were comparable to responses previously reported for the overall study population.
- EULAR response rates were higher in the CZP than the PBO treatment arm.

4.7.5.2.1. Clinical responses

ACR response

Overall, the ACR responses of the TNFi experienced population in HIKARI were numerically improved for patients receiving CZP 200 mg Q2W in comparison to those taking PBO, with ACR50 response significantly increased with CZP treatment at Week 24 (Table 29). Response rates in the TNFi experienced population were numerically similar to those reported for the overall study population, reported elsewhere.[9]

Table 29: HIKARI study: ACR response rates of the TNFi experienced monotherapy subgroup (NRI)

	PBO (n=10)	CZP 200 mg Q2W (n=6)
ACR20: n (% response rate), p, OR [95% CI]†		
Week 12	[redacted]	[redacted]
Week 24	[redacted]	[redacted]
ACR50: n (% response rate), p, OR [95% CI]†		
Week 12	[redacted]	[redacted]
Week 24	[redacted]	[redacted]
ACR70: n (% response rate), p, OR [95% CI]†		
Week 12	[redacted]	[redacted]
Week 24	[redacted]	[redacted]

FAS (NRI), †For PBO versus CZP
*p<0.05, ≠p-value not calculated.

EULAR response

The proportion of EULAR good and moderate responders in the PBO group did not change between Weeks 12 and 24 and was consistently lower than the proportion of good and moderate responders in the CZP group. Within the CZP 200 mg Q2W group, the proportion of good responders increased between Weeks 12 and 24, with a corresponding decrease in the numbers of non-responders at Week 24 (Table 30).

Table 30: HIKARI study: EULAR response rates of the TNFi experienced monotherapy subgroup (LOCF)

		PBO (n=10)	CZP 200 mg Q2W (n=6)
EULAR Response, n (%), p, OR [95% CI]†			
Week 12	Good	[redacted]	[redacted]
	Moderate	[redacted]	[redacted]
	None	[redacted]	[redacted]
Week 24	Good	[redacted]	[redacted]
	Moderate	[redacted]	[redacted]
	None	[redacted]	[redacted]

FAS (LOCF). †For PBO versus CZP.
≠p-values not calculated.

4.7.5.2.2. Disease activity

DAS28(ESR) remission

The proportion of patients in the PBO group achieving remission at Weeks 12 and 24 did not change, with [redacted] in remission at both time points. In the CZP 200 mg Q2W group, [redacted] achieved remission by Week 12 however by Week 24, [redacted] were in remission (Table 31).

Table 31: HIKARI study: DAS28(ESR) remission of the TNFi experienced monotherapy subgroup (LOCF)

		PBO (n=10)	CZP 200 mg Q2W (n=6)
DAS28 (ESR) remission, n (%), p, OR [95% CI]†			
Week 12	Remission	[redacted]	[redacted]
	Non-remission	[redacted]	[redacted]
Week 24	Remission	[redacted]	[redacted]
	Non-remission	[redacted]	[redacted]

FAS (LOCF). †For PBO versus CZP.
≠p-values not calculated

4.8 *Subgroup analysis*

No subgroup analyses have been conducted.

4.9 *Meta-analysis*

The studies used to conduct the meta-analyses were identified by the systematic review described in Section 4.1 and Table 13.

A direct meta-analysis was performed in order to pool the data from REALISTIC, J-RAPID and SWITCH for the sub-populations of patients that received CZP in combination with MTX, and MTX in the PBO arm. In the SWITCH study, patients who were secondary non-responders or intolerant to TNFi were randomised to either CZP or PBO in addition to stable background of MTX or other DMARDs. For the purpose of conducting meta-analysis, it was assumed that all patients included in the SWITCH study received concomitant MTX. A direct meta-analysis was also performed in order to pool the data from REALISTIC and HIKARI for the sub-populations of patients that received CZP monotherapy and PBO.

For all included studies, only the subgroup of patients previously exposed to a TNFi were considered, including patients from J-RAPID study (11 patients received CZP in combination with MTX and 15 patients received PBO in combination with MTX) and

patients from HIKARI study (6 patients received CZP monotherapy and 10 patients received PBO only).

Data from DOSEFLEX and PREDICT were not included in the meta-analysis. In the DOSEFLEX study, since only Week 16 ACR20 responders (after open-label treatment with CZP) were randomised at Week 18 to either one of the two dose regimens of CZP or MTX, the study design was considered not appropriate for pooling or for comparing with the data from other trials. In the PREDICT study all patients received CZP and the study did not include any common comparator.

As described in Section 4.1 and Table 8, the meta-analysis was conducted for selected outcomes (ACR20/50/70, EULAR response, and DAS28(ESR) remission at 3 months), given the data availability. Of note, the REALISTIC study and SWITCH study reported comparative data for efficacy outcomes from randomised phase at 3 months (Week 12) only. Therefore no direct meta-analysis was performed for outcomes reported at 6 months, as such outcomes were only available in J-RAPID for CZP in combination with MTX and in HIKARI for CZP monotherapy.

4.9.1 Statistical Methods

4.9.1.1 Dichotomous data

Dichotomous outcomes were summarised as the relative risk ratio (RR). The relative RR is the ratio of risks of the event in the treatment group relative to the risk of the event in the control group. Risks are defined as follows:

$$Risk = \frac{n}{N}$$

Where n represents the number of patients with the event and N represents the number of patients observed (generally the intention-to-treat (ITT) population for that treatment group).

Meta-analysis was performed in Stata statistical software. Fixed-effects estimates were calculated according to the Mantel-Haenszel (M-H) model, and random-effects estimates according to the method of DerSimonian and Laird (D-L).[78]

The results of the analysis of the RR for all the efficacy outcomes where CZP in combination with MTX compared with MTX and CZP monotherapy compared with PBO are presented as a combined forest plot. The forest plot present the effect estimate and respective confidence intervals for each study on one set of axes, along with the pooled estimate of effect. M-H represents fixed-effects results calculated using the Mantel-Haenszel model. D-L represents random-effects results calculated using the DerSimonian and Laird method.

The horizontal lines on the forest plots show the CIs; the wider the Cis the less precise the results of the trial. The vertical line corresponds to a point where treatment and control are equally as effective (1 for risk ratios; 0 for absolute risk difference and weighted mean difference). Any CI that crosses this value implies that no statistically significant effect was found in the study.

Heterogeneity between trials was explored through the calculation of I^2 values. I^2 is considered a preferred test for heterogeneity in judging consistency of evidence as it does not inherently depend on the number of studies in the meta-analyses.[79]

Thresholds for the interpretation of I^2 can be misleading, since the importance of inconsistency depends on several factors; however, a rule of thumb to the interpretation of I^2 is as follows:

- 0% to 30%: mild heterogeneity
- >30% to 60%: moderate heterogeneity
- >60% to 100%: substantial heterogeneity

The importance of the observed value of I^2 depends on (i) the magnitude and direction of effects and (ii) the strength of evidence for heterogeneity (eg. p value from the chi-squared test or CIs for I^2).[79]

4.9.2 Results

4.9.2.1 Combination with MTX

The results of the analysis of the risk ratio for all the efficacy outcomes where CZP in combination with MTX is compared with MTX are presented as a combined forest plot in Figure 42. Detailed results are presented in Appendix 8.11.1.

The estimated relative RRs from the meta-analysis indicated that patients receiving CZP in combination with MTX are three times more likely to achieve an ACR20 response and four times more likely to achieve an ACR50 response at 3 months compared to patients administered MTX alone. Similar trends were observed for the other outcomes (ACR70 response and EULAR response), with CZP in combination with MTX being associated with a higher proportion of patients achieving the clinical response compared to MTX. These results should be interpreted with caution due to the small population size of J-RAPID and SWITCH studies and heterogeneity among included studies.

4.9.2.2 Monotherapy

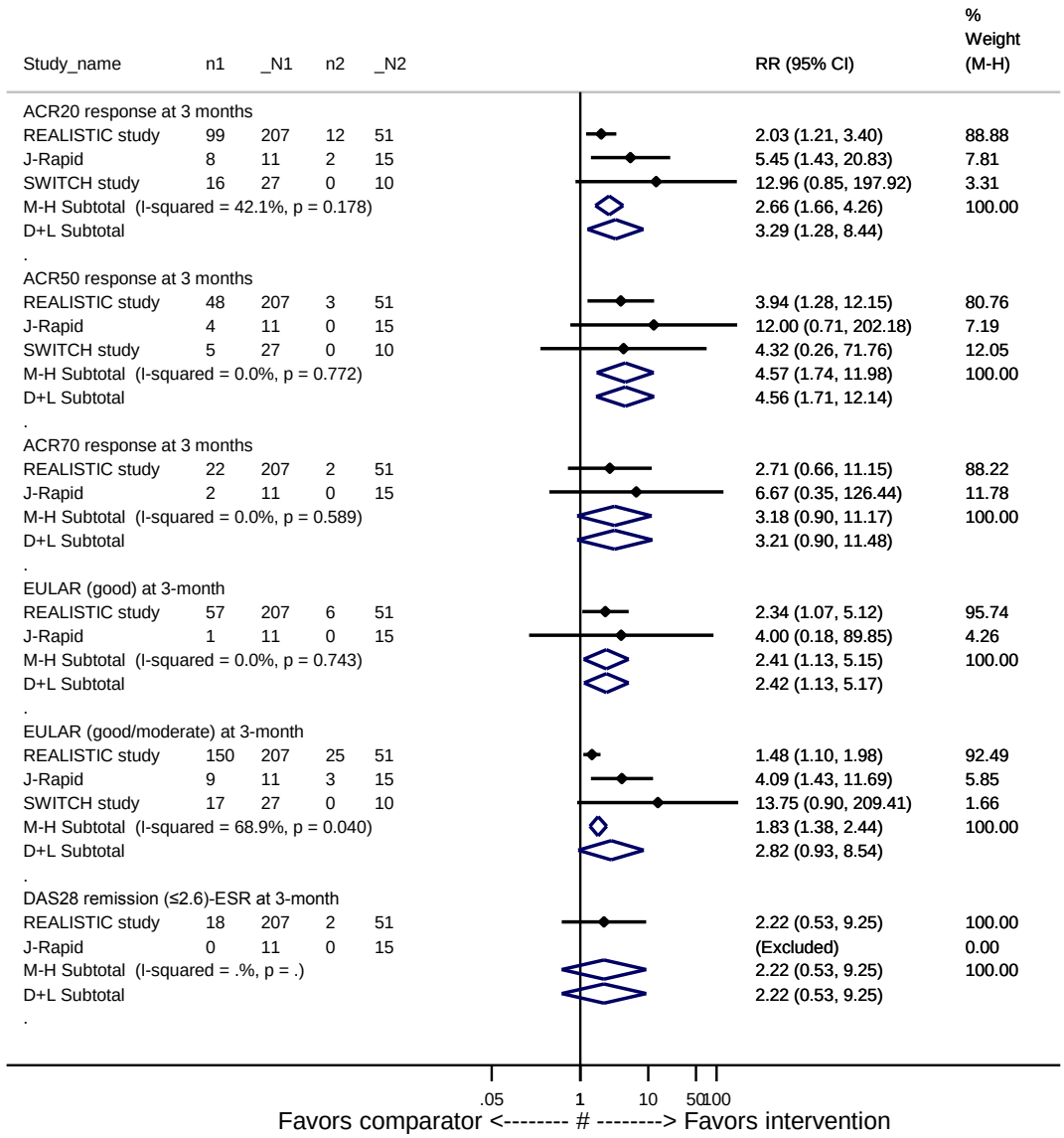
The results of the analysis of the relative RR in terms of the efficacy outcomes of CZP monotherapy compared to PBO are presented as a combined forest plot for all analysed outcomes in Figure 43. Detailed results are presented in Appendix 8.11.2.

Results estimate that patients receiving CZP alone achieved higher ACR response (ACR20/50/70) at 3 months compared to patients administered PBO. Similarly, CZP alone is associated with an increased proportion of patients achieving EULAR response and DAS28(ESR) remission compared to PBO. These results should be interpreted with caution due to the small population size for HIKARI study.

4.9.3 Heterogeneity

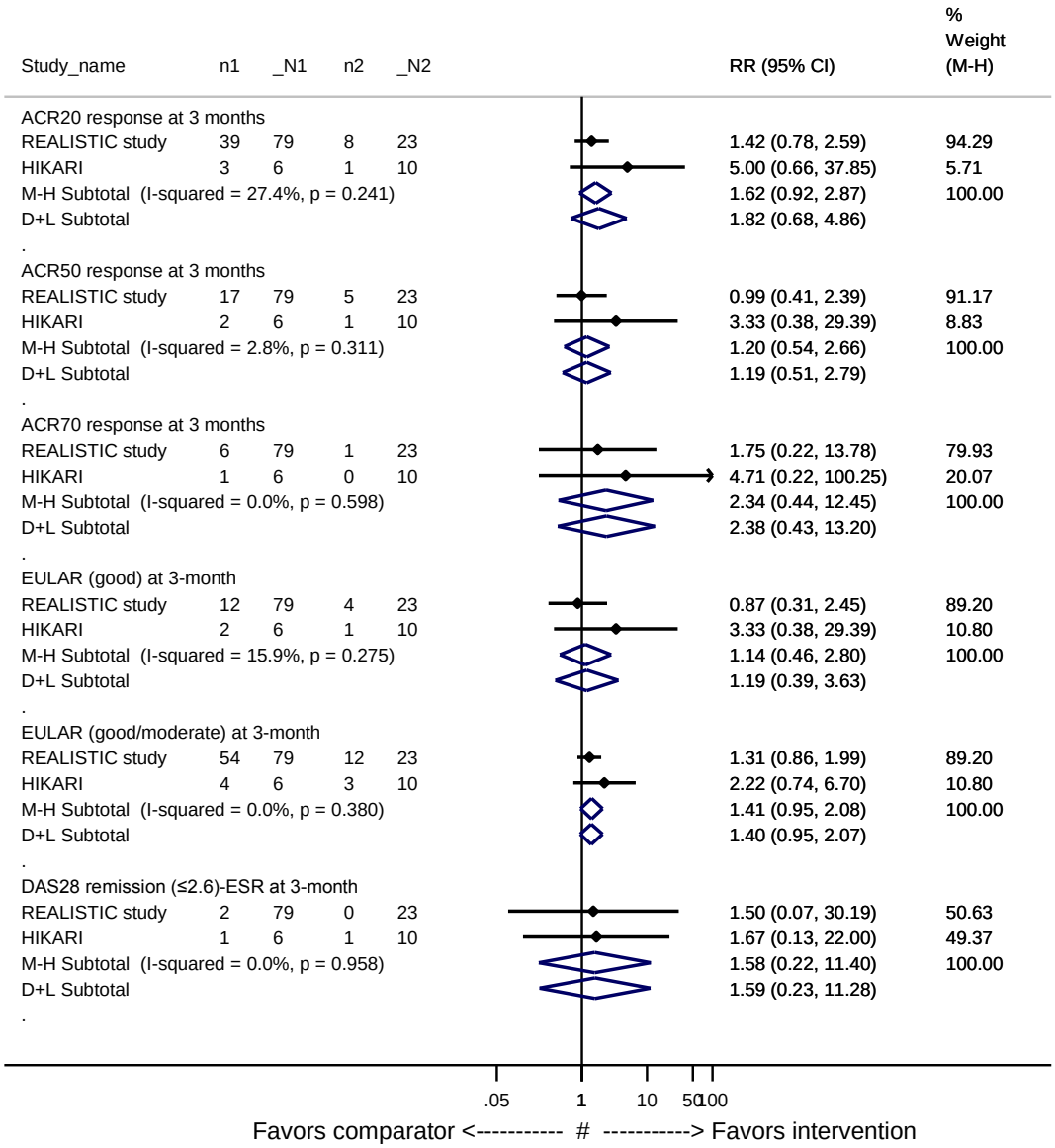
The statistical measure of heterogeneity (I^2) presented alongside meta-analysis results in Appendix 8.11.1 and 8.11.2 for both populations (combination with MTX and monotherapy, respectively) was 0% in 6 of the 11 meta-analyses presented, where data were pooled from ≥2 studies. Substantial heterogeneity was observed for EULAR good/moderate response at 3 months for the combination with MTX population (68.9%). However, direction of estimates from all the studies was aligned and the heterogeneity can be ignored. In the monotherapy population, I^2 was less than 30% for all the analysed outcomes indicating that the statistical heterogeneity among the pooled studies might not be important. In the combination with MTX population, I^2 was 42.1% for ACR20 response indicating moderate statistical heterogeneity among the pooled studies. Both fixed- and random-effects results are presented in the forest plots and meta-analyses results tables.

Figure 42: Direct meta-analysis results: all efficacy outcomes at 3 months (12 weeks), CZP + MTX versus MTX



Note: D-L represents random-effects results calculated using the DerSimonian and Laird model; M-H represents fixed-effects results calculated using the Mantel-Haenszel model;
N1 = Number of patients randomised in the CZP + MTX arm; n1 = Number of patients with response at 3-month in the CZP + MTX arm; N2 = Number of patients randomised in the MTX arm; n2 = Number of patients with response at 3-month in the MTX arm; For the direction of meta-analyses results, intervention represents CZP + MTX and comparator represents MTX alone

Figure 43: Direct Meta-analysis: All Efficacy Outcomes at Week 12 for CZP Monotherapy versus PBO



Note: D-L represents random-effects results calculated using the DerSimonian and Laird model; M-H represents fixed-effects results calculated using the Mantel-Haenszel model;
N1 = Number of patients randomised in the CZP monotherapy arm; n1 = Number of patients with response at 3 month in the CZP monotherapy arm; N2 = Number of patients randomised in the PBO arm; n2 = Number of patients with response at 3 month in the PBO arm; For the direction of meta-analyses results, intervention represents CZP monotherapy and comparator represents PBO

4.10 *Indirect and mixed treatment comparisons*

4.10.1 Search strategy

A systematic review was performed to identify all relevant RCTs for CZP and its comparators listed in the scope (ADA, ABA, ETA, IFX, CZP, GOL and TOC). Further information on the methodology of the systematic review can be found in Section 4.1.

4.10.2 Study selection

The systematic review detailed in Section 4.1 was used to identify trials relevant to the decision problem. As per the scope of this NICE submission, only trials that included at least one licensed dosing regimen were included. Some studies identified in the systematic review were excluded from the indirect treatment comparisons (ITCs) and reasons for exclusion are listed in Appendix 8.4.2.

4.10.3 Methods and outcomes of included studies

Out of the 16 studies identified by the systematic literature review, nine studies of biological DMARDs administered in combination with MTX were included in the ITC. Three of these studies provided data for CZP (REALISTIC, SWITCH, and J-RAPID), two for TOC (RADIATE and Genovese 2014), while one study each assessed GOL (GO-AFTER), ABA (ATTAIN), and RTX (REFLEX), respectively. Combe 2012 compared RTX in combination with MTX vs. ETA in combination with MTX. This trial was only included in the sensitivity analysis of the indirect comparison due to small sample size and implausible results. All trials were double-blind and conducted in multiple centres with the exception of REFLEX study which was a triple-blind study and Combe 2012, which was an open label study. Two studies that assessed CZP (REALISTIC and HIKARI) provided data for monotherapy, however, ITCs were not conducted as data were not available for any competitor treatment as monotherapy. Some studies identified in the systematic review were not included in the indirect analyses; the reasons for exclusion are listed in Appendix 8.4.2.

The study duration for the PBO-controlled randomised phase ranged from 12 weeks (REALISTIC, SWITCH, and Genovese 2014) to 26 weeks (Combe 2012 and ATTAIN), with the remaining studies with 24 weeks of double-blind, randomised phase (including J-RAPID, GO-AFTER, RADIATE and REFLEX).

Patient populations in the trials included in the indirect analyses were patients who had failed previous TNFi. All studies except Combe 2012 reported that concomitant medications were allowed.

It should be noted that in the ATTAIN study 77.8% of the included patient population received concomitant MTX while in Genovese 2014, 95% of the included patients received concomitant MTX. However, for the purpose of ITC data from these two studies were considered in the 'combination with MTX' analysis group, given that the majority of patients were on combination with MTX. Similarly, in the SWITCH study, it was assumed that all patients received concomitant MTX with other cDMARDs

All studies were conducted in patients with moderate to severe active RA and variation was observed in the included studies in terms of baseline mean HAQ-DI and disease duration.

A brief overview of the nine studies included in the indirect analysis and baseline characteristics of the patients included in these studies are presented in Appendix 8.12.1. An overview of the data for outcomes of interest available from the studies included in the ITC is provided in Appendix 8.12.2.

4.10.4 Risk of bias

A detailed critical appraisal of the studies included in indirect analysis was conducted, using the minimum criteria recommended by NICE for the quality assessment (based on Centre for Reviews and Dissemination's guidance)[80], Jadad score,[81] and allocation concealment grade (Grade A: adequate; Grade B: uncertain; Grade C: inadequate; Grade D: no allocation concealment attempted). A quality assessment of each study included in the indirect analysis is provided in the Appendix 8.12.1.3.

4.10.5 Methods of indirect treatment comparisons

ITCs were conducted by using both Bayesian network meta-analysis model (NMA)[82] and adjusted indirect treatment comparison.[83] An adjusted indirect analysis method was chosen over Bayesian for outcomes where evidence network included not more than two competing interventions. For evidence networks assessing more than two competing interventions a Bayesian NMA was performed.

Further, sensitivity analyses were conducted after inclusion of data from Combe 2012 study which randomised patients to receive RTX in combination with MTX (10 patients) or ETA in combination with MTX (10 patients) after inadequate response to ETA and RTX. Network diagrams for the feasibility of indirect treatment comparisons are presented in Appendix 8.12.3 and 8.12.4 for base case analysis and for sensitivity analyses, respectively.

4.10.5.1 Methods for Network meta-analysis

An NMA was conducted for the network of evidence where data was available to compare CZP with two or more interventions. Mixed treatment comparisons, a special case of NMA, combine direct with indirect evidence for particular pair-wise comparisons thereby synthesising a greater share of the available evidence than traditional meta-analyses.

The key to most NMA are the possibly disparate studies utilised. The mathematical handling of the studies plays a critical role in the results of the analysis. Two standard types of models could be used for conducting NMA.

The first model includes an independent parameter for each study. This analysis creates a “fixed” effect for each study and these study effects would be estimated independently. In practice this means that a distribution will be assigned to each study and these study effects will be allowed to vary independently. This model is naïve and used for comparison purposes. Due to the overlapping nature of studies, this model may yield numerically unstable results.

The second model is referred to as hierarchical model or random-effects model. This is a powerful way of treating studies arising from a common distribution of studies; hence the common link to the studies is explicitly modelled. This modelling allows understanding of variability from study to study. The hierarchical model is the standard model used in NMA and evidence syntheses as it is expected to have better performance and properties than the simple fixed effects model.

A hierarchical model is presented below:

Each study is characterized by a study parameter,

$$Study = a_S \text{ for } S=1,\ldots..NS$$

Each study is then modelled with a distribution, creating a multilevel or hierarchical approach:

$$a_S \sim N\left(m, s^2\right.$$

The population of studies is then modelled with the above distribution, which can be used to draw conclusions about possible new studies. The two parameters, μ and σ2 are then modelled as second level (hyper) prior. The analysis creates a posterior distribution of these parameters μ and σ2. The posterior also uses the population of studies to individually estimate each study, resulting in better estimates and smaller standard errors.

Both fixed effects and random effects analyses were conducted in WinBUGS with code based on that prepared by the Multi-parameter Evidence Synthesis Research Group of the Universities of Bristol and York (code available on request). This approach relies on vague prior distributions on study and treatment effects, and results in posterior distributions for relative and absolute effects. The absolute effect estimates were calculated based on PBO in combination with MTX being chosen as the principle comparator treatment. The mixed treatment comparison analysis estimated absolute effects of the principle comparator treatment (PBO in combination with MTX) by effectively pooling data from those arms of the various studies involving that treatment. Absolute effects for other treatments were derived by applying relative effects to the absolute effects estimated for the principle comparator treatment.

4.10.5.2 Methods for adjusted indirect treatment comparison

An adjusted indirect method proposed by Bucher et al. was used as the basis for the ITC for outcomes where only two interventions were compared. This method is used when g studies have compared treatment A with treatment B and h studies have compared treatment C with treatment B. The indirect estimate of association between A and C can then be calculated. The measure of treatment effect is RR.

For studies comparing treatments A and B in subgroup i the treatment effect can be calculated as a RR:

$$RR_{AB}(i) = \frac{n_A/N_A}{n_B/N_B}$$

where nA/NA = risk in the treated group and nB/NB = risk in the control group. n represents the number of patients with the event and N represents the number of patients receiving treatment (generally the ITT population for that treatment group).

If an experimental intervention has an identical effect to the control, the RR will be 1. If it reduces the chance of having the event, the RR will be less than 1; if it increases the chance of having the event, the RR will be larger than 1. The smallest value for the RR is naught when there are no events in the treated group.

Standard meta-analysis techniques (DerSimonian and Laird method) can then be used to calculate the overall effect measure for A versus B and C versus B as a weighted average of the individual effect measures of the included studies using well established command “metan” in Stata (as was done in Section 4.9 to calculate pooled estimates of CZP in combination with MTX vs MTX).

The indirect estimate of A versus C can then be calculated by taking a log transformation of the ratio of the summary RR for the direct estimates.

$$\ln(RR_{AC}) = \ln(RR_{AB}) - \ln(RR_{CB})$$

The test of this association is based on the chi-squared value for the overall association of A versus B plus the chi-squared value for the overall association of B versus C.

Programming language used to conduct indirect treatment analysis along with the methods used for derivation of missing data from the included studies are presented in Appendix 8.13.

Studies for inclusion in the indirect analysis were identified from the systematic review. Where data allows, results are presented on the clinical effectiveness of CZP compared to ABA, ETA, RTX, GOL and TOC using MTX as a common comparator for combination therapies and PBO as the common comparator for monotherapies. The measure of clinical effectiveness used in the indirect analyses is ACR20/50/70 response and EULAR response.

Indirect analyses were conducted for moderate to severe RA population only as all included studies assessed patients with moderate to severe active RA. None of the included study presented subgroup data for patients with severe disease activity at study baseline except for REALISTIC. Therefore, indirect analyses were not conducted for the severe RA population.

Results of the indirect analysis for efficacy outcomes including ACR20/50/70 response and EULAR response at 3-months and 6-months are presented in Section 4.10.6. Indirect analysis was not possible for ACR70 response at 3-months due to there being no data from the comparator studies.

Indirect comparison was also not possible for bDMARDs monotherapy due to the paucity of data from the comparator studies.

4.10.5.3 Method to analyse the heterogeneity across trials

Heterogeneity between trials of the same agents was explored through the calculation of I^2 values. The assumption is made that trials of different agents were sufficiently similar to pool without further adjustment, for example, through meta-regression. This assumption was considered to be reasonable given the identification of studies for this analysis through a systematic review with a strict inclusion/exclusion criteria.

If there is heterogeneity a random-effects approach typically assumes that true relative effects across studies are considered exchangeable (ie. the prior position of expecting underlying effects to be similar but not identical) and can be described as a sample from a normal distribution the mean of is the pooled relative effect and whose standard deviation (SD) reflects the heterogeneity. The standard error obtained from a fixed-effect analysis will be too small if there is heterogeneity between trials (beyond random

variation). A general test for model fit (standard model fitting diagnostics, DIC (Deviance Information Criteria) will be applied to assess and capture sensitivity to assumptions (random-effect or fixed-effect).[84] Model selection was based on the DIC model with lower DIC and residual deviance value indicating a better fit. Results from both models are presented.

In the NMA analysis, for each outcome one common heterogeneity parameter tau^2 was assumed across comparisons. The parameter tau^2 corresponds to the variance of underlying distribution. For each tau^2, we will estimate 95% credible intervals as well. Tau^2 value≥1 indicates that there is intra-study variability and vice-versa, though there is no specific range for this measure.[85]

A simple exploration of heterogeneity was considered sufficient, as the value of any further complex analyses would be limited by the relatively small amount of data provided by the available trials. For example, a meta-regression of how results varied by baseline HAQ score would be desirable, but the results would not be robust given the small number of trials. Incorporating covariates in a meta-regression beyond study and treatment effect could lead to the overfitting of sparse data.

4.10.6 Indirect treatment comparison results

4.10.6.1 Basecase results

The indirect analysis conducted showed that CZP in combination with MTX is at least as effective to the other comparators considered in all of the cases in patients with moderate to severe RA that were previously exposed to TNFi. The wide credible intervals noted in most of the analyses results reflect the minimal differences in relative clinical effect between CZP and the comparators considered. These results should be interpreted with caution due to heterogeneity in the patient population among the included studies. Model fit characteristics from NMA are presented in Appendix 8.12.5.

4.10.6.1.1. ACR20 response at 3 months

There were five studies which contributed data for this analysis (Table 32).

No significant difference was observed when CZP in combination with MTX was compared with TOC in combination with MTX and ABA in combination with MTX (Figure 44). These results should be interpreted with caution due to the small population size of SWITCH and J-RAPID and heterogeneity among the included studies.

Table 32: ACR20 response at 3-months

Study name	Interventions compared	ITT patients	Patient with response
REALISTIC study	CZP + MTX	[redacted]	[redacted]
	MTX	[redacted]	[redacted]
J-RAPID study	CZP + MTX	[redacted]	[redacted]
	MTX	[redacted]	[redacted]
SWITCH study	CZP + MTX	27	16
	MTX	10	0
ATTAIN study	ABA + MTX	258	118
	MTX	133	24
Genovese 2014	TOC + MTX	43	27
	MTX	22	7

Figure 44: Forest plot for ITC for ACR20 response at 3 months: RR with 95%CrIs

4.10.6.1.2. ACR50 response at 3 months

There were four studies which contributed data for this analysis (Table 33).

No significant difference was observed between CZP in combination with MTX and TOC in combination with MTX (Table 34). These results should be interpreted with caution

due to the small population size of SWITCH and J-RAPID and heterogeneity among the included studies.

Table 33: ACR50 response at 3 months

Study name	Interventions compared	ITT patients	Patient with response
REALISTIC study	CZP + MTX	[redacted]	[redacted]
	MTX	[redacted]	[redacted]
J-RAPID study	CZP + MTX	[redacted]	[redacted]
	MTX	[redacted]	[redacted]
SWITCH study	CZP + MTX	27	5
	MTX	10	0
Genovese 2014	TOC + MTX	43	10
	MTX	22	0

Table 34: Results of ITC for ACR50 response at 3 months: RRs with 95% CIs

Intervention	Comparator	RR	LCI	UCI
CZP + MTX	[redacted]	[redacted]	[redacted]	[redacted]

4.10.6.1.3. EULAR (good/moderate) response at 3 months

There were four studies which contributed data to this analysis (Table 35).

No significant difference was observed between CZP in combination with MTX and RTX in combination with MTX (Table 36). These results should be interpreted with caution due to the small population size of SWITCH and J-RAPID and heterogeneity among the included studies.

Table 35: EULAR (good/moderate) response at 3 months

Study name	Interventions compared	ITT Patients	Patient with response
REALISTIC study	CZP + MTX	[redacted]	[redacted]
	MTX	[redacted]	[redacted]
J-RAPID study	CZP + MTX	[redacted]	[redacted]
	MTX	[redacted]	[redacted]
SWITCH study	CZP + MTX	27	17
	MTX	10	0
REFLEX study	RTX + MTX	311	204
	MTX	209	68

EULAR = European League Against Rheumatism; ITT = Intention-To-Treat; MTX = Methotrexate

Table 36: Results of ITC for EULAR (good/moderate) response at 3 months: RRs with 95% CIs

Intervention	Comparator	RR	LCI	UCI
CZP + MTX	[redacted]	[redacted]	[redacted]	[redacted]

4.10.6.1.4. EULAR (good) response at 3 months

There were three studies which contributed data to this analysis (Table 37).

No significant difference was observed between CZP in combination with MTX and RTX in combination with MTX (Table 38). These results should be interpreted with caution due to the limited number of studies and the small population size of J-RAPID.

Table 37: EULAR (good) response at 3 months

Study name	Interventions compared	ITT Patients	Patient with response
REALISTIC study	CZP + MTX	[redacted]	[redacted]
	MTX	[redacted]	[redacted]
J-RAPID study	CZP + MTX	[redacted]	[redacted]
	MTX	[redacted]	[redacted]
REFLEX study	RTX + MTX	311	33
	MTX	209	10

Table 38: Results of ITC for EULAR (good) response at 3 months: RRs with 95% CIs

Intervention	Comparator	RR	LCI	UCI
CZP + MTX	[redacted]	[redacted]	[redacted]	[redacted]

4.10.6.1.5. ACR20 response at 6 months

There were five studies which contributed data to this analysis (Table 39).

Results were in favour of CZP in combination with MTX when compared with ABA, GOL, RTX, and TOC, all in combination with MTX, but were not statistically significant (Figure 45). These results should be interpreted with caution due to the small population size of J-RAPID and heterogeneity among the included studies.

Table 39: ACR20 response at 6 months

Study name	Interventions compared	ITT Patients	Patient with response
J-RAPID study	CZP + MTX	[redacted]	[redacted]
	MTX	[redacted]	[redacted]
REFLEX study	RTX + MTX	311	152

Study name	Interventions compared	ITT Patients	Patient with response
	MTX	209	36
RADIATE study	TOC + MTX	175	85
	MTX	161	16
ATTAIN study	ABA + MTX	258	129
	MTX	133	26
GO-AFTER study	GOL + MTX	101	36
	MTX	107	15

Figure 45: Forest plot for ITC for ACR20 response at 6 months: RRs with 95%CrIs

4.10.6.1.6. ACR50 response at 6 months

There were five studies which contributed data to this analysis (Table 40).

Results were in favour of CZP in combination with MTX when compared with ABA, GOL, RTX, and TOC, all in combination with MTX, but were not statistically significant (Figure 46). These results should be interpreted with caution due to the small population size of J-RAPID and heterogeneity among the included studies.

Table 40: ACR50 response at 6 months

Study name	Interventions compared	ITT Patients	Patient with response
J-RAPID study	CZP + MTX	[redacted]	[redacted]
	MTX	[redacted]	[redacted]
REFLEX study	RTX + MTX	311	80
	MTX	209	10
RADIATE study	TOC + MTX	175	49
	MTX	161	6
ATTAIN study	ABA + MTX	258	52
	MTX	133	5
GO-AFTER study	GOL + MTX	101	20
	MTX	107	4

Figure 46: Forest plot for ITC for ACR50 response at 6 months: RRs with 95%CrIs

4.10.6.1.7. ACR70 response at 6 months

There were five studies which contributed data for this analysis (Table 41).

Results were in favour of CZP in combination with MTX when compared with ABA, GOL, RTX, and TOC, all in combination with MTX, but were not statistically significant (Figure 47). These results should be interpreted with caution due to the small population size of J-RAPID and heterogeneity among the included studies.

Table 41: ACR70 response at 6 months

Study name	Interventions compared	ITT Patients	Patient with response
J-RAPID study	CZP + MTX	[redacted]	[redacted]
	MTX	[redacted]	[redacted]
REFLEX study	RTX + MTX	311	36
	MTX	209	2
RADIATE study	TOC + MTX	175	21
	MTX	161	2
ATTAIN study	ABA + MTX	258	26
	MTX	133	2
GO-AFTER study	GOL + MTX	101	12
	MTX	107	3

Figure 47: Forest plot for ITC for ACR70 response at 6-months: RRs with 95%CrIs

4.10.6.1.8. EULAR (good/moderate) response at 6 months

There were five studies which contributed data for this analysis (Table 42).

CZP in combination with MTX appeared to be associated with a better EULAR (good/moderate) response rate than ABA, GOL, RTX, and TOC, all in combination with MTX, however, results were not statistically significant using the random effect model (Figure 48). These results should be interpreted with caution due to the small population size of J-RAPID and heterogeneity among the the included studies.

Table 42: EULAR (good/moderate) response at 6 months

Study name	Interventions compared	ITT Patients	Patient with response
J-RAPID study	CZP + MTX	[redacted]	[redacted]
	MTX	[redacted]	[redacted]
REFLEX study	RTX + MTX	311	194
	MTX	209	44
ATTAIN study	ABA + MTX	258	130
	MTX	133	33
RADIATE study	TOC + MTX	175	115
	MTX	161	26
GO-AFTER study	GOL + MTX	101	49
	MTX	107	27

Figure 48: Forest plot for ITC for EULAR (good/moderate) response at 6 months: RRs with 95%CrIs

4.10.6.1.9. EULAR (good) response at 6 months

There were three studies which contributed data to this analysis (Table 43).

CZP in combination with MTX appeared to be associated with a better EULAR (good) response rate than ABA or RTX in combination with MTX, however, results were not statistically significant (Figure 49). These results should be interpreted with caution due to the small population size of J-RAPID.

Table 43: EULAR (good) response at 6 months

Study name	Interventions compared	ITT Patients	Patient with response
J-RAPID study	CZP + MTX	[redacted]	[redacted]
	MTX	[redacted]	[redacted]
REFLEX study	RTX + MTX	311	45
	MTX	209	4
ATTAIN study	ABA + MTX	258	30

Study name	Interventions compared	ITT Patients	Patient with response
	MTX	133	4

Figure 49: Forest plot for ITC for EULAR (good) response at 6 months: RRs with 95%CrIs

4.10.6.2 Sensitivity analysis results

Sensitivity analyses were also conducted for moderate to severe RA population using data from Combe 2012. Sensitivity analyses were conducted for EULAR response only at both 3-months and 6 months due to limited availability of data. Results of the sensitivity analyses are presented below.

4.10.6.2.1. EULAR (good/moderate) response at 3 months

There were five studies which contributed data for this analysis (Table 44).

No significant difference was observed when CZP in combination with MTX was compared with RTX or ETA in combination with MTX (Figure 50). These results should be interpreted with caution due to the small population size of SWITCH, J-RAPID and Combe 2012 study and heterogeneity among the included studies.

Table 44: EULAR (good/moderate) at 3 months

Study name	Interventions compared	ITT Patients	Patient with response
REALISTIC study	CZP + MTX	[redacted]	[redacted]
	MTX	[redacted]	[redacted]
J-RAPID study	CZP + MTX	[redacted]	[redacted]
	MTX	[redacted]	[redacted]
SWITCH study	CZP + MTX	27	17
	MTX	10	0
Combe 2012	RTX + MTX	10	5
	ETA + MTX	10	7
REFLEX study	RTX + MTX	311	204
	MTX	209	68

Figure 50: Forest plot for ITC for EULAR (good/moderate) response at 3 months: RRs with 95% CrIs

4.10.6.2.2. EULAR (good) response at 3-months (sensitivity analysis)

There were four studies which contributed data for this analysis (Table 45).

No significant difference was observed when CZP in combination with MTX was compared with RTX or ETA in combination with MTX (Figure 51). These results should be interpreted with caution due to the small population size of J-RAPID and Combe 2012 study.

Table 45: EULAR (good) response at 3 months

Study name	Interventions compared	ITT Patients	Patient with response
REALISTIC study	CZP + MTX	[redacted]	[redacted]
	MTX	[redacted]	[redacted]
J-RAPID study	CZP + MTX	[redacted]	[redacted]
	MTX	[redacted]	[redacted]
REFLEX study	RTX + MTX	311	33
	MTX	209	10
Combe 2012	RTX + MTX	10	2
	ETA + MTX	10	3

Figure 51: Forest plot for ITC for EULAR (good) response at 3 months: RRs with 95% CrIs

4.10.6.2.3. EULAR (good/moderate) response at 6-months (sensitivity analysis)

There were six studies which contributed data to this analysis (Table 46).

CZP in combination with MTX appeared to be associated with a better EULAR (good/moderate) response rate than ABA, ETA, GOL, RTX, and TOC, all in combination with MTX, however, results were not statistically significant using the random effect model (Figure 52). These results should be interpreted with caution due to the small population size of J-RAPID and Combe 2012 study and heterogeneity among the included studies.

Table 46: EULAR (good/moderate) response at 6 months

Study name	Interventions compared	ITT Patients	Patient with response
J-RAPID study	CZP + MTX	[redacted]	[redacted]
	MTX	[redacted]	[redacted]
REFLEX study	RTX + MTX	311	194
	MTX	209	44
Combe 2012	RTX + MTX	10	5
	ETA + MTX	10	8
ATTAIN study	ABA + MTX	258	130
	MTX	133	33
RADIATE study	TOC + MTX	175	115
	MTX	161	26
GO-AFTER study	GOL + MTX	101	49
	MTX	107	27

Figure 52: Forest plot for ITC for EULAR (good/moderate) response at 6 months: RRs with 95% CrIs

4.10.6.2.4. EULAR (good) response at 6-months (sensitivity analysis)

There were four studies which contributed data to this analysis (Table 47).

CZP in combination with MTX appeared to be associated with a better EULAR (good) response rate than ABA, ETA, and RTX, all in combination with MTX, however, results were not statistically significant (Figure 53). These results should be interpreted with caution due to the small population size of J-RAPID and Combe 2012 study.

Table 47: EULAR (good) response at 6 months

Study name	Interventions compared	ITT Patients	Patient with response
J-RAPID study	CZP + MTX	[redacted]	[redacted]
	MTX	[redacted]	[redacted]
REFLEX study	RTX + MTX	311	45
	MTX	209	4

Study name	Interventions compared	ITT Patients	Patient with response
Combe 2012	RTX + MTX	10	2
	ETA + MTX	10	3
ATTAIN study	ABA + MTX	258	30
	MTX	133	4

Figure 53: Forest plot for ITC for EULAR (good) response at 6 months: RR with 95% CIs

4.11 *Non-randomised and non-controlled evidence*

4.11.1 Non-randomised and non-controlled trials considered

In addition to the RCT evidence, one relevant non RCT was identified (Table 48). The ARTIS study is an observational, real-world investigation using the Swedish Rheumatology Quality Register (SRQ) database.

Table 48: List of relevant non-randomised and non-controlled evidence

Study acronym	Objective	Population	Intervention	Comparator	Primary study reference	Justification for inclusion in the submission
ARTIS	To investigate the effectiveness and survival on CZP in a real-world setting in relation to disease activity at baseline	Patients diagnosed with RA between October 2009 and June 2013, as recorded on the SRQ	CZP, as per license dose	None	Chatzidio-nysiou et al. 2015[67]	Presented efficacy data for CZP in a patient population who had failed at least one TNFi

4.11.2 Methodology of non-randomised and non-controlled trials

Data for patients who were diagnosed with RA and started CZP treatment within the study period (1st October 2009–31st June 2013) were analysed. Efficacy data including DAS28 and HAQ scores, DAS28 remission and EULAR response were collected at 3 and 6 months after initiation of CZP therapy.[67] Further details of the methodology of the ARTIS study are provided in Table 49, and outcomes reported by the trial are presented in Table 50.

Table 49: Non-RCT trial methodology

	ARTIS[67]	
Location	Sweden	
Trial design	Observational, registry based study using the SRQ	
Eligibility criteria for participants	Inclusion: • RA patients recorded in SRQ who initiated CZP therapy between 1st October 2009 and 31st June 2013	Exclusion: • RA patients recorded in SRQ who did not receive CZP therapy between 1st October 2009 and 31st June 2013
Settings and locations where the data were collected	SRQ database	
Intervention (n=) and comparator(s) (n=)	• CZP (n=945)	
Primary outcomes (including scoring methods and timings of assessments)	• DAS28 and HAQ score change from baseline at 3 and 6 months	
Secondary/tertiary outcomes (including scoring methods and timings of assessments)	• Proportion of patients with DAS28 remission at 3 and 6 months • Proportion of patients with EULAR response at 3 and 6 months • Survival on drug at end of follow-up	
Preplanned- subgroups	Stratification by number of prior biologic TNFi agents: • 0 • 1 • ≥2 Stratification by disease activity: • High activity (DAS28 >5.1) • Other activity (DAS28 ≤5.1)	

Table 50: Overview of non-RCTs and outcomes considered

Trial	Outcome	Data source
ARTIS	• DAS28 response and score • HAQ score • EULAR response • Survival on drug	• Chatzidionysiou 2015[67]

4.11.3 Statistical analysis of non-randomised and non-controlled trials

A total of 945 patients were initially considered in the study. Three subgroups were formed from this initial cohort, (1) TNFi-naïve (n=540), (2) prior exposure to 1 TNFi (n=215), and (3) prior exposure to 2 or more TNFi agents (n=190). TNFi treatments were previously discontinued for various reasons (ineffectiveness, intolerance, other).[67]

The mean DAS28, change in DAS28, HAQ, and change in HAQ scores at the 3 and 6 month time-point were compared across the groups by ANOVA followed by the Bonferroni test for post-hoc comparisons between the groups. The level of statistical significance was set to 5%. All analyses were performed for the whole cohort and stratified by number of TNFis previously discontinued.[67]

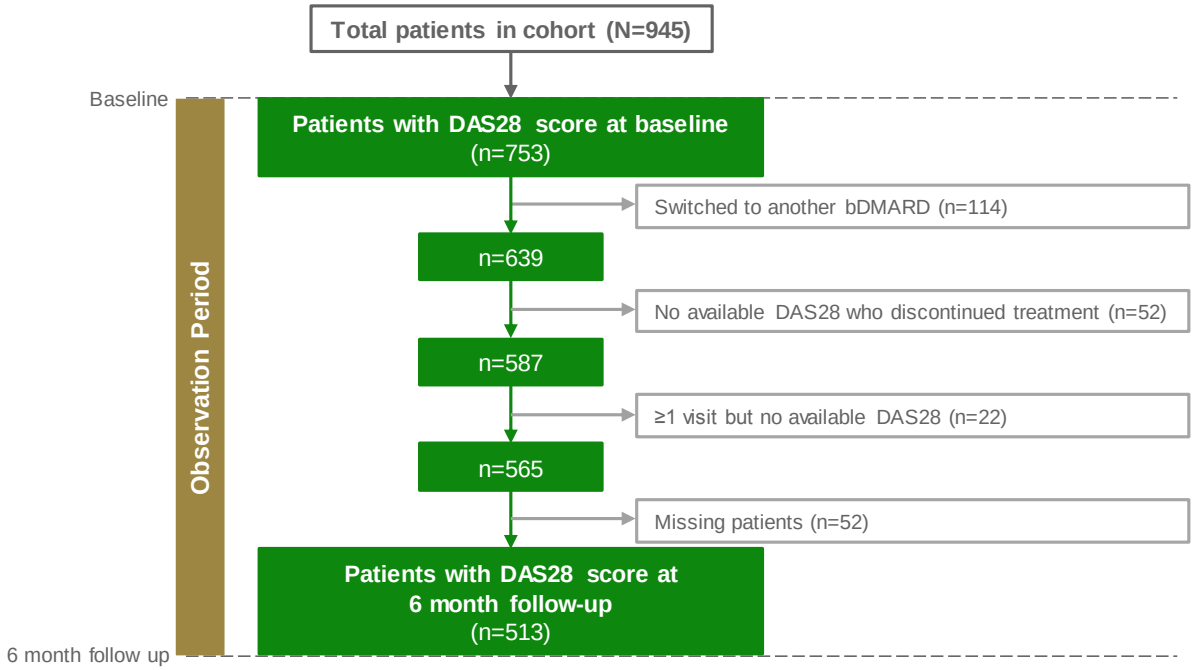
The observational design of the ARTIS study gave rise to certain limitations in the statistical analyses and sources of bias. Baseline characteristics between groups under comparison were not completely balanced and significant differences were observed, which introduced a risk for confounding. The authors tried to partially overcome this problem by adjusting for variables that differed significantly between groups in a Cox regression analysis. Missingness was another problem typical of register-based observational studies and was explored in the ARTIS study; 70% of patients who had available DAS28 at baseline also had available DAS28 information at 6 months. From the total number of patients with no available DAS28 at 6 months, 166 were imputed as non-responders as they had available information on treatment discontinuation (114 switched to another biological treatment before entry into the 3-month window and 52 discontinued treatment with CZP) and 74 had missing information.[67]

However, there were substantial strengths to the ARTIS study, such as the large number of patients included in the cohort and reflective of clinical practice, the opportunity to examine the effectiveness of treatment in both the TNFi-naïve population and in patients who had already failed one or more TNFis. This study is therefore the first observational study to examine the effectiveness of CZP in this context.[67]

4.11.4 Participant flow of non-randomised and non-controlled trials

A diagram presenting the flow of participants enrolled into the full randomised set in the ARTIS trial is presented in Figure 54. A total of 953 patients from the SRQ with RA who initiated treatment with CZP were considered in the analysis. Of these, 753 patients had DAS28 scores and baseline, and 513 had DAS28 scores at the 6 months follow-up. The most common reason for not having DAS28 scores at 6 months was due to the patient switching to an alternative TNFi.[67]

Figure 54: Flow of patients in ARTIS analysis



Adapted from Chatzidionysiou et al. (2015)[67]

The baseline characteristics of the patients included in ARTIS are presented in Table 51. Patients had on average over 9 years disease duration and the majority were taking cDMARDs at baseline (65.4%).

Given the observational design of the study, there were some imbalances of characteristics between the subgroups stratified by prior TNFi exposure:[67]

- The proportion of females was significantly greater in the group with ≥2 prior TNFis compared to the TNFi-naïve group (p=0.002)
- The mean disease duration was significantly greater with increasing prior TNFi exposure, between all subgroups (p<0.0001 between all groups)
- The proportion of patients taking cDMARDs was significantly lower with prior TNFi exposure, between all subgroups (p=0.08 and p<0.0001 for TNFi-naïve vs

1 prior TNFi and ≥2 prior TNFis, respectively, and p=0.04 for 1 prior TNFi vs ≥2 prior TNFi)

Table 51: ARTIS study: characteristics of participants across different groups

	Overall cohort (N=945)	TNFi-naïve (n=540)	1 prior TNFi (n=215)	≥2 prior TNFis (n=190)
Mean age (SD), years	56.4 (13.8)	55.7 (13.9)	57.7 (13.7)	57.1 (13.6)
Female, n (%)*	75.2%	72.2%	75.3%	83.7%
Mean disease duration, years (SD)**	[937] 9.1 (3.6–17.7)	[535] 6 (2–12.8)	[213] 10.9 (5.6–18.9)	[188] 15 (9.9–23.7)
Use of DMARDs, % Yes***	65.4%	70.2%	63.7%	53.7%
DAS28(ESR) score, mean (SD)	[753] 4.6 (1.4)	[447] 4.6 (1.4)	[159] 4.6 (1.4)	[147] 5.0 (1.5)****
HAQ score, mean (SD)	[820] 1.1 (0.7)	[474] 1.0 (0.6)	[181] 1.1 (0.6)	[165] 1.4 (0.7)*****

SD: standard deviation;
*p=0.002 (0 prior TNFi vs 2 prior TNFi); p=0.04 (1 prior TNFi vs 2 prior TNFi)
**p< 0.0001 between all groups pairwise
***p=0.08 (0 prior TNFi vs 1 prior TNFi); p<0.0001 (0 prior TNFi vs 2 prior TNFi); p=0.04 (1 prior TNFi vs 2 prior TNFi)
****p=0.01 vs TNFi-naïve and p=0.04 vs 1 prior TNFi
*****p<0.0001 vs TNFi-naïve and p=0.003 vs 1 prior TNFi
N numbers for group presented in square brackets where they differ from the column heading
Adapted from Chatzidionysiou et al. (2015)[67]

4.11.5 Quality assessment of non-randomised and non-controlled trials

The full quality assessment of ARTIS is presented in Appendix 8.10. Overall, the publication reported the outcomes, patient characteristics, confounders and findings in sufficient detail. The external validity of the study was good as the cohort was selected from a real-world registry. However the internal validity of this study was reduced due to the inherent bias that resulted from a lack of blinding and comparators, while the statistical analyses and subgroup analyses were clearly stated. The confounding/selection bias that is inherent to non-randomised trials was also present. Of note, the ARTIS trial reports a rate of survival on CZP which was similar to that reported in other observational studies.[67]

4.11.6 Clinical effectiveness results of the relevant non-randomised and non-controlled evidence

Summary of CZP clinical effectiveness in the TNFi experienced population of ARTIS[67]

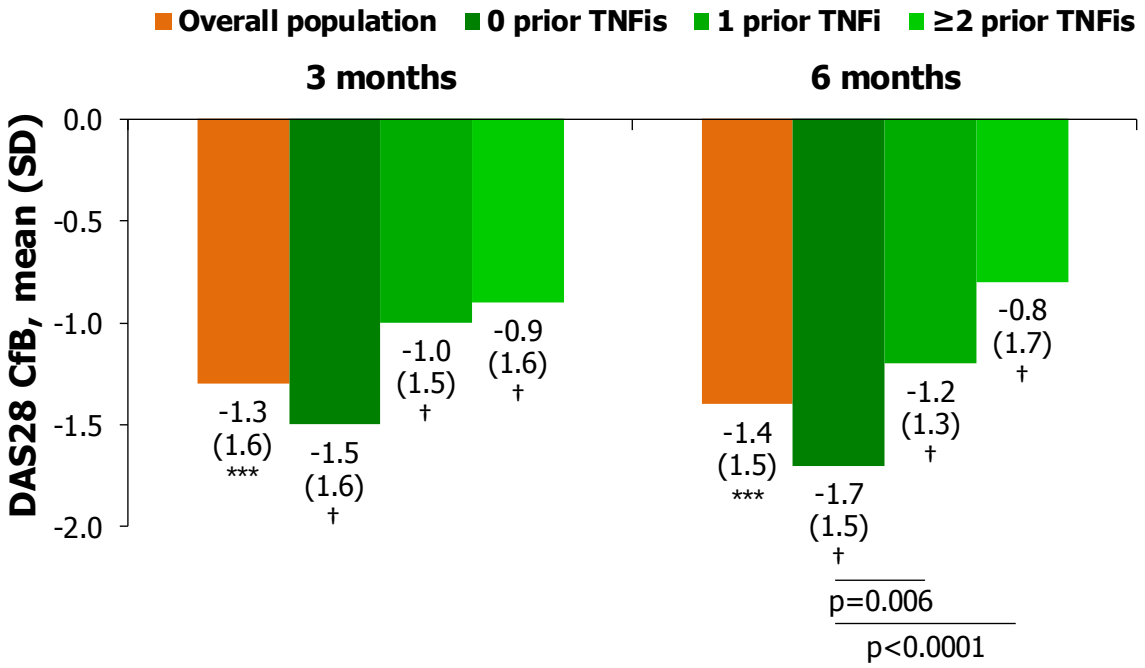
- Significant reductions in disease activity and impact on physical function were observed for both TNFi naïve and experienced patients at 3 and 6 months following CZP treatment.
- Overall, changes from baseline disease activity, as measured by DAS28 scores, were significantly greater in TNFi naïve patients, while improvements in physical function were consistent across all patient groups.

4.11.6.1 Disease activity (DAS28 scores)

On average, significant reductions in disease activity (DAS28(ESR) scores)[3] were seen in the whole cohort after 3 and 6 months' treatment with CZP (p<0.0001 at both time-points) (Figure 55).[67]

For the subgroups exposed to 0, 1 and ≥2 prior TNFis, the mean baseline DAS28 score was significantly higher in patients with ≥2 prior TNFis than those with 1 or 0 prior TNFis (p=0.04 and p=0.01, respectively) (Table 51). Following treatment with CZP, each of the subgroups achieved significant improvements in DAS28 scores at 3 and 6 months (p value not presented); at 6 months, changes from baseline in DAS28 were significantly greater in TNFi naïve patients compared to those exposed to 1 or ≥2 prior TNFis (p=0.006 and p<0.0001, respectively) (Figure 55).[67]

Figure 55: ARTIS study: mean change from baseline in DAS28 scores after 3 and 6 months treatment with CZP (NRI)‡



***p<0.0001

[3]Components utilised for score are listed as TJC, SJC, general health and ESR by Chatzidionysiou et al.

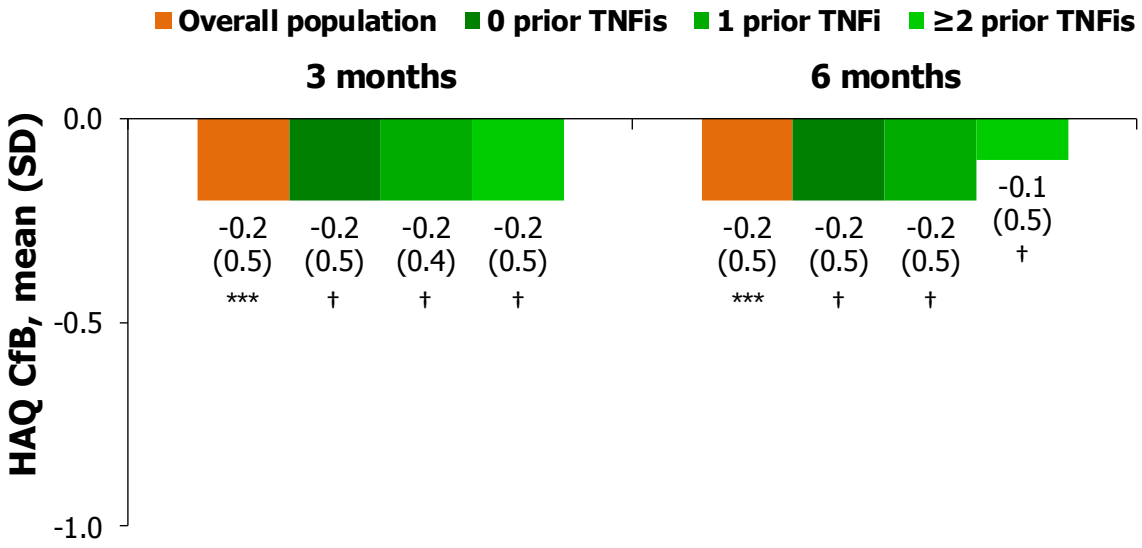
†significant reduction; p-value not reported
‡166 patients with missing follow-up data were imputed as non-responders at 6 months based on available information on treatment discontinuation.
Group sizes at 3 months were n=321, n=197, n=67 and n=57, respectively
Group sizes at 6 months were n=440, n=267, n=89 and n=84, respectively
Adapted from Chatzidionysiou et al. (2015)[67]

4.11.6.2 Impact on physical function (HAQ)

Significant improvements in physical function (ie reductions in HAQ scores) were reported in the whole cohort after 3 and 6 months' treatment with CZP (p<0.0001 at both time-points) (Figure 56).[67]

For the subgroups exposed to 0, 1 and ≥2 prior TNFis, the mean baseline HAQ score was significantly higher in patients with ≥2 prior TNFis than those with 1 or 0 prior TNFis (p=0.003 and p<0.0001, respectively) (Table 34). Following treatment with CZP, each of the subgroups achieved significant improvements in physical function change from baseline of HAQ scores at 3 and 6 months (p value not presented); the change from baseline was comparable between all subgroups at both time-points (not significant [ns] for all comparisons) (Figure 56).[67]

Figure 56: ARTIS study: mean change from baseline in HAQ scores after 3 and 6 months treatment with CZP (Censoring)‡



***p<0.0001
†significant reduction; p-value not reported
‡Patients who had no follow-up visit were censored.
Group sizes at 3 months were n=382, n=229, n=86 and n=67, respectively
Group sizes at 6 months were n=472, n=273, n=100 and n=93, respectively
Adapted from Chatzidionysiou et al. (2015)[67]

4.12 *Adverse reactions*

The safety of CZP has been evaluated in a large number of studies in patients with moderate to severe RA. In this submission, the safety data emerging from the four above mentioned RCTs (REALISTIC, DOSEFLEX, PREDICT and SWITCH) are presented, with additional safety data from REALISTIC from the TNFi experienced population. Additionally, a pooled analysis has been conducted that evaluates the safety of CZP across a number of RCTs and OLEs in RA.[10] This section presents these individual, and overall pooled safety evaluations.

4.12.1 Adverse reactions in REALISTIC, DOSEFLEX, PREDICT, SWITCH

4.12.1.1 REALISTIC

4.12.1.1.1. Overall trial population

The overall incidence of AEs amongst the entire trial population was comparable between the CZP and PBO groups in the double-blind phase of the REALISTIC study (67.5% vs 61.7%, respectively; Table 52).[4] The majority of AEs in both groups were of mild to moderate intensity. Upper respiratory tract infections, nausea, headaches and flare of RA were the AEs most commonly reported (Table 52).[4] Injection and infusion-site reactions occurred in a greater proportion of CZP patients than PBO patients (5.8% vs. 1.0%, respectively; Table 52). SAEs were reported in 6.1% of patients in the CZP group and 5.7% of patients in the PBO group during the double-blind phase of the study (Table 52). The most common SAEs were infections occurring in 22 (2.6%) CZP patients and 4 (1.9%) PBO patients. There were no reported cases of tuberculosis (TB) in either group.[4]

In the OLE population from Weeks 12 to 28, similar incidences of AEs were reported in the CZP→CZP and PBO→CZP groups (67.7% vs. 77.2%, respectively; Table 53).[64] SAEs were experienced by 11.4% of patients in the PBO→CZP group and 7.3% of patients in the CZP→CZP group during the OLE phase, the most common of which were serious infections (Table 53).[64]

Two deaths occurred in the CZP group up to Week 12, both of which were ruled as possibly related to CZP: one case of sigmoid diverticulitis in a 73-year-old man with pancreatitis, which occurred 56 days after first CZP dose, and one of necrotizing pneumonia, which occurred 20 days after the first CZP dose in a 63-year-old man with unstable diabetes who was treated with corticosteroids and refused hospitalisation when the pneumonia was diagnosed at the emergency room department.[4] In the OLE up to Week 28, one additional death occurred in the CZP→CZP group, which was due

to small-cell lung cancer, and one additional death occurred in the PBO→CZP, which was due to myocardial infarction.[64]

Table 52: Safety up to the end of the 12 week double-blind phase of the REALISTIC study (overall safety population)

Exposure and AEs	CZP 200 mg Q2W (n=846)	PBO (n=209)
Duration of exposure, PY	196.4	48.9
Any AEs by maximum intensity, n (%)		
Mild	248 (29.3)	56 (26.8)
Moderate	257 (30.4)	58 (27.8)
Severe	66 (7.8)	15 (7.2)
In AEs, incidence rate/100 PY (n, patient %)		
Any AEs[a]	522.1 (571, 67.5)	483.2 (129, 61.7)
Infection and infestations	143.9 (245, 29.0)	112.5 (48, 23.0)
Upper respiratory tract infections	59.3 (112, 13.2)	41.5 (19, 9.1)
Headaches NEC	24.2 (47, 5.6)	23.5 (11, 5.3)
Nausea and vomiting symptoms	21.5 (42, 5.0)	28.2 (13, 6.2)
Rheumatoid arthropathies	18.8 (37, 4.4)	37.0 (17, 8.1)
SAEs[b]	26.7 (52, 6.1)	25.8 (12, 5.7)
Serious infections	11.1 (22, 2.6)	8.3 (4, 1.9)
Lower respiratory tract and lung infections	3.5 (7, 0.8)	2.1 (1, 0.5)
Streptococcal infections	0 (0, 0)	2.1 (1, 0.5)
Urinary tract infections	2.5 (5, 0.6)	4.2 (2, 1.0)
Death	1.0 (2, 0.2)	0 (0, 0)
AEs leading to withdrawal	20.6 (40, 4.7)	17.1 (8, 3.8)
Injection and infusion site reactions	25.3 (49, 5.8)	4.2 (2, 1.0)

AE: adverse event; CZP: certolizumab pegol; PBO: placebo; NEC: not elsewhere classified; PY: patient years; SAE: serious adverse event.
Bold: system organ class; Italics: high level term
[a]AEs occurring in >5.0% of patients in either treatment group are presented below the column subheading "Any AEs" in the table. [b]Any important medical event including events that do not require hospitalisation, such as certain opportunistic infections.
Adapted from Weinblatt et al. (2012)[4]

Table 53: Safety in the OLE period of the REALISTIC study (Week 12 to Week 28; overall safety population)†

Exposure and AEs	CZP 200 mg Q2W→ CZP 200 mg Q2W n=770	Week 12 PBO→CZP 200 mg Q2W n=184
Any AEs, incidence/100 PY (n, %)	239.1 (521, 67.7)	328.9 (142, 77.2)
SAEs[a], incidence/100 PY (n, %)	13.0 (56, 7.3)	20.6 (21, 11.4)
Serious infections, incidence/100 PY (n, %)	4.1 (18, 2.3)	5.7 (6, 3.3)
Death, n (%)	1 (0.5)	1 (0.1)
AEs leading to withdrawal, n (%)	3 (1.6)	26 (3.4)
Injection and infusion site reactions, incidence/100 PY (n, %)	8.8 (9, 4.9)	3.2 (14, 1.8)

AE: adverse event; CZP: certolizumab pegol; PBO: placebo; PY: patient years; SAE: serious adverse event.
†Patients who completed 12 weeks of treatment with either CZP 200 mg Q2W or PBO during the double-blind phase entered the OL phase and subsequently received active treatment (CZP 200 mg Q2W) for ≥16 weeks
[a]Any important medical event including events that do not require hospitalisation, such as certain opportunistic infections.
Adapted from Weinblatt et al. (2015)[64]

4.12.1.1.2. Patients with prior TNFi experience

In the population from REALISTIC who had experience of prior TNFi agents, the incidence of AEs up to the Week 12 double-blind phase in the CZP arm (68.1%) was similar to that in the overall safety population (67.5%) (Table 54).[86] The incidence of AEs in the PBO group with prior TNFi use was 50.0% (Table 54).[86] SAEs were reported in 7.9% of patients in the CZP group with prior TNFi use and 5.0% of patients in the PBO group with prior TNFi use during the double-blind phase of the study (Table 54).[86]

Table 54: Safety up to the end of the 12 week double-blind phase of the REALISTIC study (patients with prior TNFi use)

Exposure and AEs	CZP 200 mg Q2W (n=317)	PBO (n=80)
AEs, n (%)		
Any AEs	216 (68.1)	40 (50.0)
Infections and infestations	93 (29.3)	19 (23.8)
SAEs[a]	25 (7.9)	4 (5.0)
AEs leading to withdrawal	17 (5.4)	2 (2.5)

AE: adverse event; CZP: certolizumab pegol; PBO: placebo; SAE: serious adverse event.
Bold: system organ class
[a]Any important medical event including events that do not require hospitalisation, such as certain opportunistic infections.
Adapted from Weinblatt et al. (2011)[86]

4.12.1.2 DOSEFLEX

4.12.1.2.1. Overall trial population

During the DOSEFLEX trial, all patients were initially exposed to CZP 200 mg Q2W for 16 weeks during the run-in phase, before being randomised to their treatment groups for a further 18 weeks of double-blind study. The safety data for all patients across the run-in phase are presented in Table 55. At least one treatment-emergent AE (TEAE) was reported in 76% of all participants in the run-in phase.

Safety data were also recorded during the double-blind phase, after the initial 16 weeks of CZP exposure in all groups (Table 56). The overall incidence of AEs amongst the entire DOSEFLEX trial population was comparable between both CZP treatment groups (200 mg Q2W and 400 mg Q4W) and the PBO group in the double-blind phase of the DOSEFLEX study (62.9% vs 60.9% vs 62.3%, respectively; Table 56). The most common AEs in the PBO, CZP 200 mg Q2W, and CZP 400 mg Q4W groups were in the systems of infections and infestations, musculoskeletal and connective tissue disorders, gastrointestinal disorders, and respiratory, thoracic and mediastinal disorders. Among these, there were no relevant differences between groups, except for the respiratory, thoracic, and mediastinal disorders (eg. cough) where the PBO group was subject to

more AEs. There were no deaths, opportunistic infections, TB infections, or malignancies reported (Table 56).[5]

There were no SAEs reported in the PBO group during the double-blind phase as compared to 5 patients (7.1%) and 2 patients (2.9%) in the CZP 200 mg Q2W and CZP 400 mg Q4W groups, respectively. The most common SAEs were infections and infestations; 3 serious infections occurred in the CZP 200 mg Q2W and no serious infections occurred in the CZP 400 mg Q4W group. There were no instances of injection site pain and 1 instance of a local injection site rash, reported in the CZP 200 mg Q2W group during the double-blind phase (Table 56).[5]

No safety analyses have been conducted specifically in the TNFi experienced population of the DOSEFLEX trial.

Table 55: Overall summary of safety during the 16 week run-in phase of the DOSEFLEX study (modified enrolled set)

AEs, n (%)	Overall run-in, phase (n=333)
Any TEAE	[redacted]
Severe[b] TEAEs	[redacted]
Serious TEAEs	[redacted]
Drug-related[a] TEAEs	[redacted]
Serious and drug-related TEAEs	[redacted]
Discontinuation due to TEAEs	[redacted]
Permanent study drug discontinuation	[redacted]
Temporary study drug discontinuation	[redacted]
Deaths	[redacted]

[redacted]

Source: UCB data on file

Table 56: Safety during the double-blind phase of the DOSEFLEX study (overall safety set)[a]

AEs	CZP 200 mg Q2W (n=70)	CZP 400 mg Q4W (n=70)	PBO† (n=69)
AEs, incidence rate/100 PY (n, patient %)			
Any AEs[b]	312.1 (44, 62.9)	299.9 (42, 60.9)	323.6 (43, 62.3)
Infection and infestations	104.9 (20, 28.6)	132.4 (25, 36.2)	136.2 (24, 34.8)
Upper respiratory tract infections	23 (5, 7.1)	36.2 (8, 11.6)	46.5 (10, 14.5)
Nasopharyngitis	4.4 (1, 1.4)	4.4 (1, 1.4)	18.4 (4, 5.8)
Sinusitis	9 (2, 2.9)	13.1 (3, 4.3)	0
Urinary tract infection	23.1 (5, 7.1)	27.6 (6, 8.7)	33.4 (7, 10.1)
Ear infections	0	13.3 (3, 4.3)	0
Musculoskeletal/connective tissue disorders	37.6 (8, 11.4)	51.4 (11, 15.9)	64.2 (13, 18.8)
Arthralgia	4.5 (1, 1.4)	22.5 (5, 7.2)	8.9 (2, 2.9)
Back pain	13.5 (3, 4.3)	0	4.4 (1, 1.4)
RA aggravation	4.4 (1, 1.4)	8.9 (2, 2.9)	27.7 (6, 8.7)
Pain in extremity	8.9 (2, 2.9)	0	13.5 (3, 4.3)
Nervous system disorders	22.8 (5, 7.1)	17.8 (4, 5.8)	4.4 (1, 1.4)
Dizziness	13.5 (3, 4.3)	0	4.4 (1, 1.4)
Headache	9 (2, 2.9)	0	0
Skin/subcutaneous tissue disorders	22.7 (5, 7.1)	22.7 (5, 7.2)	22.4 (5, 7.2)
Rash	8.9 (2, 2.9)	0	4.4 (1, 1.4)
Respiratory/thoracic/mediastinal disorders	28 (6, 8.6)	4.4 (1, 1.4)	46.9 (10, 14.5)
Cough	0	0	13.4 (3, 4.3)
Gastrointestinal disorders	43.9 (9, 12.9)	37.9 (8, 11.6)	41.6 (9, 13)
Nausea	13.8 (3, 4.3)	0	4.4 (1, 1.4)
General disorders/administration site conditions	27.8 (6, 8.6)	13.3 (3, 4.3)	22.8 (5, 7.2)
Pyrexia	18.1 (4, 5.7)	0	4.4 (1, 1.4)
SAEs[c]	23.1 (5, 7.1)	8.8 (2, 2.9)	0
Serious infections	13.6 (3, 4.3)	0	0
Cardiac disorders	4.5 (1, 1.4)	0	0
Musculoskeletal/connective tissue disorders	9 (2, 2.9)	4.4 (1, 1.4)	0
Respiratory/thoracic/mediastinal disorders	0	0	0
AE leading to death	0	0	0
AEs leading to withdrawal[c]	58.4 (12, 17.1)	27.5 (6, 8.7)	37.3 (8, 11.6)
AEs leading to permanent discontinuation	18.4 (4, 5.7)	4.4 (1, 1.4)	0

AE: adverse event; CZP: certolizumab pegol; PBO: placebo; PY: patient years; SAE: serious adverse event.
Bold: system organ class; Italics: high level term; Underline: preferred term
†Patients in the PBO arm were treated with CZP 200 mg Q2W for 16 weeks during the run-in phase
[a]Safety set relates to all randomised patients, not including one patient who did not receive any treatment; [b]Any AEs occurring in >3% patients; [c]Temporary and permanent discontinuations
Adapted from Furst et al. (2015)[5]

4.12.1.3 PREDICT

The tolerability and safety of CZP therapy in patients with moderate to severe RA throughout 52 weeks of therapy were comparable between treatment arms. AEs were reported in 76.0% of patients (559 of 736), and SAEs occurred in fewer than 10% (71 of

736) of patients (Table 57).[6] Two deaths were reported: one due to acute myocardial infarction and one due to septic shock following pneumonia.[87]

No safety analyses have been conducted specifically in the TNFi experienced population of the PREDICT trial.

Table 57: Safety during the 52 weeks of the PREDICT study (safety set)

AEs	All Patients (n=736)	RAPID3-assigned (n=369)	CDAI-assigned (n=367)
[redacted]	[redacted]	[redacted]	[redacted]
Events, n (patient %) [#][a]			
All AEs	559 (76.0) [2,145]	270 (73.2) [1,070]	289 (78.7) [1,075]
Severe AEs	76 (10.3) [120]	37 (10.0) [65]	39 (10.6) [55]
SAEs	71 (9.6) [112]	32 (8.7) [50]	39 (10.6) [62]
Discontinuations due to AEs	78 (10.6) [112]	32 (8.7) [49]	46 (12.5) [63]
Drug-related AEs[b]	173 (23.5) [325]	88 (23.8) [156]	85 (23.2) [169]
AEs leading to death	2 (0.3) [6]	0	2 (0.5) [6]

AE: adverse event.
[a]Number of individual AE occurrences; [b]AEs with relationship of 'related' or those with missing responses
Adapted from Curtis et al. (2015)[6]

4.12.1.4 SWITCH

During the 24 weeks of study in the SWITCH trial, AEs occurred in 59.3% and 40.0% of patients in the CZP and PBO groups, respectively (Table 58). The TEAEs in both treatment groups were mild (43.8% CZP; 75.0% PBO) or moderate (56.3% CZP; 25.0% PBO), as no severe events occurred. The most frequent AEs were upper respiratory infections (17.6% CZP; 0% PBO), cough (7.4% CZP; 10% PBO), or headache (7.4% CZP; 0% PBO). One SAE of gastrointestinal bleeding (considered unrelated to CZP) occurred in the open-label phase. No opportunistic infections, TB or serious infections occurred, as did no deaths.[7]

Table 58: Safety during the 24 week SWITCH study[a]

Exposure and AEs	CZP 200 mg Q2W + cDMARDs→ CZP 200 mg Q2W + cDMARDs (n=27)	PBO + cDMARDs→ CZP 200 mg Q2W + cDMARDs (n=10)[b]
AEs, n (%)		
Any AEs	16 (59.3)	4 (40.0)
Mild	7 (25.9)	3 (30.0)
Moderate	9 (33.3)	1 (10)
Severe	0 (0)	0 (0)
Related AEs	3 (11.1)	0 (0)
SAEs[b]	0 (0)	0 (0.0)
AEs leading to withdrawal	0 (0)	1 (10.0)
Death due to AEs	0 (0)	0 (0)

AE: adverse event; CZP: certolizumab pegol; PBO: placebo; SAE: serious adverse event.
[a]Including the 12 week double-blind phase and 12 week open label phase; [b]PBO patients commenced treatment with CZP during the 12 week open label phase
Adapted from Schiff et al. (2014)[7]

4.12.2 Pooled analysis

Data from 10 completed RCTs of CZP in RA and several OLEs were pooled across all doses, accumulated as of 30th November 2011. Reported AE occurred between the first dose and 84 days after the last dose. All deaths, SIEs and malignancies were reviewed by external experts, classified according to predefined rules, and validated by an external steering committee. Incidence rates (IRs) and event rates (ERs) per 100 PY are presented.[10]

4,049 RA patients who received CZP were included in the safety pooling; total exposure 9,277 PY, mean exposure 2.1 years (range 0.04–7.6). Of this analysis, 744 (18.4%) of patients had previous exposure to a TNFi or other bDMARD.

SIEs, most frequently pneumonia (IR 0.73/100 PY), were the most common SAE, occurring more frequently in CZP compared to PBO-treated patients in RCTs (IR 5.61/100 PY vs 1.35/100 PY, odds ratio (OR) 4.35, 95% CI 0.65 to 29.30). SIE rates were lower in the CZP-treated population including OLEs (ER 4.33/100 PY). 44 patients developed TB (IR 0.47/100 PY), 39 from high endemic regions.[10] This includes participants recruited into earlier trials of RA, before recommendations were made to increase the stringency of TB screening and treatment of any latent TB infections. Following the introduction of such measures, and the raised awareness of TB, the incidence of TB has reduced.[88] 58 deaths occurred in CZP-exposed patients (IR 0.63/100 PY) and the ER for malignancy was 0.78/100 PY.[10]

The authors concluded that no new or unexpected safety signals associated with CZP emerged in this updated long-term safety analysis and, while SIE rates were higher for CZP than for PBO in RCTs, the rate decreased with continued exposure to CZP. These rates are consistent with data previously reported for CZP and other TNFi treatments.[10]

4.12.3 Additional studies reporting safety data

In addition to the safety data from the included studies and pooled analysis presented above, several other studies also provide safety data that is relevant to the populations considered by this submission.

4.12.3.1 Yun et al. (2015) – Real world evidence of bDMARD switching

Information is limited regarding the comparative safety of bDMARDs in older RA patients (therefore with more co-morbidities), and for more recently approved biologic agents including CZP, GOL and TOC. Real-world health data can provide direct comparisons of treatment outcomes in routine care settings; the work of Yun et al. sought to determine if the risk of hospitalised infection among US Medicare RA patients

differed between specific biologics (ETA, ADA, CZP, GOL, IFX, ABA, RTX and TOC). The authors conducted a retrospective cohort study of 2006–2011 Medicare claims data for RA beneficiaries from the Centres for Medicare and Medicaid Services Chronic Condition Data Warehouse. Only bDMARD switchers were included; RA patients who started a new course of ADA, CZP, ETA, GOL, IFX, ABA, RTX or TOC during the analysis period, after having been treated with a different biologic agent in the 12-month period prior to baseline were considered in the analysis. Follow up started at the time of initiation of a new biologic and ended at the earliest date of: hospitalised infection; 12 months after biologic initiation; ≥30 day gap in exposure; switch to another biologic; death; loss of Medicare coverage; malignancy; other autoimmune disease (PsA, psoriasis, AS or inflammatory bowel disease); or end of study (31st December 2011). The outcome was the first hospitalised infection, defined as an inpatient hospital discharge diagnosis of any infection (bacterial, viral and opportunistic infections), as coded by hospital admission reason. Medical records were not available to confirm infections. Instead, an algorithm was used to identify infection based on physician discharge diagnosis codes (positive predictive value ≥80%).[12]

The study by Yun et al. demonstrated that Medicare RA patients who had previously been exposed to different biologics had a significantly higher one-year risk of hospitalised infection when subsequently exposed to ETA, IFX or RTX compared to ABA (HR [95% CI]; 1.24 [1.07, 1.45], 1.39 [1.21, 1.60], 1.36 [1.21, 1.53], respectively), as presented in Table 59. Subgroup analyses showed that in high risk patients (those patients with a higher than median overall infection risk score), patients subsequently exposed to ETA, IFX or RTX had a higher one-year risk of hospitalised infection compared to ABA (HR [95% CI]; 1.43 [1.07, 1.91], 1.41 [1.07, 1.86], 1.33 [1.03, 1.73]). In low risk patients (those patients with a lower than median overall infection risk score), only patients subsequently exposed to RTX and IFX had a higher one-year risk of hospitalised infection compared to ABA (HR [95% CI]; 1.40 [1.21, 1.62], 1.35 [1.13, 1.61], respectively). The risk of hospitalised infection compared to ABA did not increase significantly when either high risk (patients with an infection risk score greater than the median risk score) or low to average risk (those patients with a lower than baseline infection risk score) patients were switched to CZP (HR [95% CI]; 1.39 [0.95, 2.02], 0.94 [0.72, 1.24], respectively).[12]

Table 59: Crude incidence (unadjusted) and adjusted risk (HR) of first hospitalised infection

	n	Events[a]	PYs	Crude IR/100 PYs (95% CI)	Crude absolute risk difference	Adjusted HR[b] (95% CI)
ADA	4845	317	2171	14.6 (13.1, 16.3)	0.015	1.08 (0.93, 1.25)
CZP	1866	106	747	14.2 (11.7, 17.2)	0.011	1.07 (0.86, 1.32)
GOL	3814	87	616	14.1 (11.5, 17.4)	0.010	1.14 (0.90, 1.44)
ETA	1394	275	1726	15.9 (14.2, 17.9)	0.028	1.24 (1.07, 1.45)*
IFX	3944	370	2178	17.0 (15.3, 18.8)	0.039	1.39 (1.21, 1.60)*
RTX	4718	541	2898	18.7 (17.2, 20.3)	0.056	1.36 (1.21, 1.53)*
TOC	2016	129	863	14.9 (12.6, 17.8)	0.018	1.10 (0.89, 1.34)
ABA	9204	705	5377	13.1 (12.2, 14.1)	Reference	Reference

*Considered significant by the authors
[a]First hospitalised infection during follow-up; [b]Adjusted for infection risk score, number of previous biologics used, disability status, glucocorticoid use at baseline, MTX use at baseline, most recent biologic prior to baseline and Medicaid eligibility; ABA was used as the reference group.
Adapted from Yun H et al. (2015)[12]

4.12.3.2 Simard et al. (2011) – Safety of bDMARDs in bDMARD-experienced RA

Simard et al. identified all adult patients with RA (n = 9,612), PsA (n = 1,417) and other SpA (n = 1,652) initiating a first bDMARD therapy between 1st January 1999 and 31st December 2008, registered in the Swedish Biologics Register (ARTIS) to assess the extent to which contemporary patients who start or switch bDMARD therapies are comparable with those patients who gave rise to the currently available data on effectiveness and safety.[11]

A total of 3,121 patients with RA, 385 with SpA and 371 with PsA started treatment with a second or third bDMARD during the study period. The median age was relatively unchanged for first, second and third starters with RA, but the proportion of female patients was higher for second and third-line bDMARD initiators compared to first-line initiators (79 and 80% vs 76%, respectively).[11]

Disease activity at treatment initiation, as measured by DAS28, appeared slightly higher for second- and third-line starters compared to first-line starters. The overall proportion of patients with RA with a history of hospitalisation with infection increased from 12% at first start to 20 and 24% at second and third bDMARD treatment start, respectively. History of infection at second treatment start was greater among those who switched to a non-TNFi regimen compared with those initiating a second TNFi therapy (32 vs 18%). Results were nearly identical when the analysis was restricted to second-line therapy initiators from 2005 onwards (non-TNFi, n=242; TNFi, n=1,548) to account for differences in drug availability in the earlier calendar period (but the gap for history of infection widened, 36 vs 19%).[11]

The data from this study are real-world registry data, in comparison to clinical trials of which typically recruit a more homogenous population of patients that may be of overall lower risk of infections compared to the general population. This registry study highlights that the risk of infection may increase with subsequent exposure to alternative biologics.

4.12.3.3 Curtis et al. (2015) – Safety data of CZP treatment in bDMARD-naïve RA

Additional data on the safety of CZP has been presented by Curtis et al. These authors conducted an analysis of bDMARD-naïve patients from the RAPID-1 and RAPID-2 RCTs to examine whether improvement in disease activity over time is associated with a reduced risk of SIEs in CZP-treated RA patients. Post-hoc analyses included all SIEs with onset after the first CZP dose and up to 84 days after the last CZP dose or withdrawal. Patients were categorised by decrease in DAS28(CRP) between specific time points and the assessment performed before the first CZP dose (absolute decrease: <0.6; 0.6–1.2; 1.2–2.6; ≥2.6); pts with >1 DAS28(CRP) assessment in a time interval were categorised using the largest decrease in that interval.[89]

In total, 1,506 bDMARD naïve CZP patients were included, with a total exposure of 5,778.6 PY (median exposure per patient: 4.79 PY). In total, 201 patients reported ≥1 SIE (IR 3.66/100 PY [3.17–4.21]) over ~6 years of CZP treatment. According to a multivariate Cox proportional hazards model, patients experiencing a flare had more than double the risk of SIEs (HR 2.21 [1.41–3.47]); also, an increment of 0.125 in HAQ-DI was associated with a 27% increase in the risk of SIEs (HR 1.27 [1.03–1.56]). By contrast, a one unit decrease in DAS28(CRP) was associated with a 17% reduction in the risk of SIEs (HR 0.83 [0.75–0.92]). Furthermore, the observed incidence of SIEs over time was generally lower among patients with an absolute decrease ≥2.6 in DAS28(CRP). This study highlights that over time, lower disease activity was associated with a reduced risk of SIEs in CZP-treated RA patients.[89]

4.13 *Interpretation of clinical effectiveness and safety evidence*

4.13.1 Findings of the clinical evidence in this submission

The efficacy of CZP in patients with moderate to severe RA with previous exposure to TNFi treatment is supported by six RCTs, which have all provided consistently favourable results. Across the TNFi exposed populations within REALISTIC, DOSEFLEX, PREDICT, SWITCH, J-RAPID and HIKARI, CZP proved relevant, clinically meaningful benefits across a broad spectrum of clinical and patient-relevant outcomes, including improvements in signs and symptoms, reductions in disease activity, fatigue and sleep problems, improvements in physical function, health related QoL and

workplace and household productivity. Such effects have also been maintained in the longer term, suggesting that the positive effects of CZP can result in long-term treatment benefits. Additionally, CZP was effective both as a monotherapy or in combination with MTX, with similar improvements to those seen in the overall population of patients previously exposed to TNFi. Comparable treatment responses were found in 200 mg Q2W and 400 mg Q4W dosing arms of CZP in the DOSEFLEX. These results are consistent with previously reported data in patients treated with CZP, regardless of prior TNFi exposure, highlighting the similar benefit of CZP in a broad population.

The safety profile of CZP in RA has been comprehensively evaluated in a pooled analysis based on 10 completed RCTs and their OLEs providing data on a total of 4,049 CZP-treated patients with moderate to severe disease activity, collectively comprising 9,277 PY, with individual patient exposures of up to 7.6 years.[10] The strength of such an analysis is that similarities in clinical trials allow for a pooling of data and a more comprehensive safety assessment. In this longer-term safety analysis for CZP in RA, no new safety signals have emerged, and the AE profile of CZP, including SAEs and SIEs, was in line with that previously reported and that of other biologics while the risk benefit profile of CZP remains favourable.[10] These conclusions are also supported by real-life data and post-marketing safety monitoring, including pharmacovigilance data, which UCB has submitted regularly to the EMA.[11, 12]

4.13.2 Strengths and limitations of the clinical evidence in this submission

This submission presents data from a range of high-quality RCTs with populations relevant to this submission. The largest breadth of evidence comes from the REALISTIC study, which included a broader, clinically relevant population of RA patients, including those with prior TNFi use and a range of concomitant DMARDs, including monotherapy, which more closely resembles the mix of patients seen in daily clinical practice.[4] This increases the external validity of these results as well as the applicability of conclusions from REALISTIC to the decision problem.

Three of the RCTs, REALISTIC, DOSEFLEX and PREDICT, recruited a high proportion of patients with prior TNFi exposure, providing robust pre-defined subgroups for presentation in this submission that match the patient population defined in the decision problem. Alongside this, one trial, SWITCH, only recruited participants with previous TNFi exposure, being directly relevant to this submission. In addition, the RCTs provide evidence on a broad spectrum of clinical and patient relevant outcomes, thus providing evidence on the holistic treatment approach with CZP. The favourable evidence for CZP in the treatment of RA patients with prior TNFi exposure is further supported by the real-world findings of the large-scale ARTIS registry.

A number of limitations exist for the evidence presented here. In the REALISTIC study, the sample size was planned to consist of 419 patients in order to achieve 90% power, however only 400 patients were recruited, thus leaving the trial slightly underpowered.[4] The internal validity of the REALISTIC study is supported by its adequate randomisation process and comparable baseline characteristics between the treatment groups.

In the DOSEFLEX study, post-hoc analyses of the patients stratified by TNFi experience which led to a small sample size, for which statistical inferential analyses could not be conducted. This limits the reliability with which conclusions can be drawn from these results. Additionally, all patients in DOSEFLEX received CZP 200 mg Q2W in combination with MTX during the run-in phase, therefore this study only provides comparative evidence of CZP in combination with MTX versus PBO from the run-in phase onwards. The internal validity of the study is supported by the comparability of treatment groups at baseline and the patient population examined in this submission is directly relevant to the population stated in the decision problem.

One other limitation is the low number of patients previously exposed to a TNFi in J-RAPID and HIKARI.[8, 9] The SWITCH study had a small patient population as well, due to premature cessation of enrolment, although the primary endpoint was reached by a significant proportion of patients at this point in time. However, as the study was designed to enrol 102 subjects to reach 80% power, but after early termination, just 37 patients were enrolled, this resulted in a substantial reduction in power. In addition to this, very limited information regarding the statistical analyses that were conducted to compare the primary endpoint between treatment groups was provided in the manuscript, further limiting the reliability of conclusions drawn from these data.[7] Additionally, although the randomisation and blinding processes implemented in this study were not reported in detail, the comparability of patient characteristics between the two treatment arms was high, which supports the internal validity of this study. The SWITCH study also recruited patients treated with mixed cDMARDs (MTX or others) however the data for the specific subgroup of patients with concomitant MTX at baseline was not available in the manuscript and so has not been assessed in this submission.

Although the PREDICT study provides high quality data about the efficacy of CZP in a population relevant to this submission, of which it is well aligned with other studies investigating CZP, the trial is limited by the lack of a comparator group (ie PBO),[6] leading to its exclusion from the ITC. Similarly, the internal validity of the ARTIS study was reduced due to the inherent bias from lack of comparators, as well as lack of blinding,[67] which also resulted in exclusion of the latter study from the ITC.

With no limit on the number of prior TNFi therapies in ARTIS, DOSEFLEX and PREDICT but limitation to no more than two in REALISTIC, J-RAPID and HIKARI, the

impact on clinical effectiveness in patients exposed to numerous prior TNFi therapies has not been examined sufficiently here and requires further investigation.

All studies included in this analysis presented data in patients directly relevant to the decision problem and the external validity of these studies is supported by the use of licensed doses of CZP, as described in the SmPC.

Finally, no clinical evidence is available directly comparing CZP with other TNFis in their effectiveness for treating active RA in TNFi experienced patients.

CZP for the treatment of RA patients with prior TNFi exposure should not be considered as an end of life treatment.

4.14 *Ongoing studies*

There are no ongoing studies of relevance to this submission.

5 Cost effectiveness

The results of the cost effectiveness analysis strongly support the case that CZP is a cost effective treatment option in TNFi-IR patients for whom RTX is contraindicated or withdrawn or for whom RTX therapy cannot be given because MTX is contraindicated or withdrawn.

In these populations, CZP either dominates existing therapies or is cost-effective within the bounds of conventional thresholds (ie. <<£30,000).

Population A (adults previously treated with other DMARDs including at least one TNFi)

- Basecase: Treatment with CZP+MTX followed by RTX+MTX and standard therapies is more effective (+0.288 QALY gained per patient) and more expensive (+£9,938 per patient) than a sequence of RTX+MTX followed by standard therapies (deterministic ICER of £34,516 per QALY gained). Probabilistic sensitivity analyses showed that CZP+MTX had a low probability (3.0%) of being cost-effective at the £20,000 per QALY threshold but increased to 37% for a threshold of £30,000 per QALY.

Population B (adults for whom RTX is contraindicated or withdrawn)

- Basecase: The full incremental analysis showed that at WTP thresholds of between £3,641 and £129,319, CZP+MTX is the optimal treatment strategy, compared to all alternative biologics considered. Probabilistic sensitivity analyses showed that CZP+MTX had a high likelihood of at least 95% of being cost-effective at thresholds of both £20,000 and £30,000 per QALY.

Population C (adults for whom RTX therapy cannot be given because MTX is contraindicated or withdrawn)

- Basecase: The full incremental analysis showed that at thresholds of between £4,985 and £123,915, CZP monotherapy is the optimal treatment strategy. Probabilistic sensitivity analyses showed that CZP monotherapy was very close to 100% probability of being cost-effective at thresholds of both £20,000 and £30,000 per QALY.

5.1 *Published cost effectiveness studies*

5.1.1 Identification of studies

A systematic literature review was performed to identify relevant evidence of the costs and costs-effectiveness of drug therapies in patients previously exposed to TNFis. Published economic evaluations and cost and resource use studies were identified from a search of the following electronic databases (date of search: 3 November 2015): EMBASE, MEDLINE and Cochrane.

The searches were limited to evidence published in the last 10 years (ie. after 2005). This was justified on the basis that a previous review by the West Midlands Health Technology Assessment Collaboration conducted as part of the multiple technology appraisal TA195 (ADA, ETA, IFX, RTX and ABA for the treatment of RA after the failure of a TNFi) had identified four relevant publications in this patient group, all of which were published after 2008. This previous review therefore validates the assumption that

the 10-year search restriction would not exclude relevant evaluations for ADA, ETA, IFX, RTX and ABA.

For CZP, GOL, and TOC, it is likely that any published economic evaluations would have been reported within the time frame for the search (ie. between 2005 and 2015) given that the public drug acquisition price for these therapies became available after 2005, and following European regulatory approval.

Therefore, despite the 10-year search restriction, it is expected that the database search will have identified all published economic evaluations that are considered relevant to this appraisal. The full list of search terms is provided in Appendix 8.14.

The inclusion criteria for the review are summarised in Table 60, and are based on the criteria used by the West Midlands Health Technology Assessment Collaboration in their review of economic evaluations for TA195.[90] The original inclusion criteria were expanded to include CZP, GOL, and TOC as interventions of interest, alongside other biologic therapies. Only English language publications/abstracts were considered.

Table 60: Study inclusion criteria used

Inclusion criteria	Population: Adults with active RA who have had an inadequate response to a TNFi. Interventions: ADA, ETA, IFX, RTX, ABA, CZP, TOC, GOL Outcomes: Cost effectiveness/utility, cost estimates, quality of life estimates. Study design: Cost-consequence/benefit analyses; cost effectiveness /utility analyses; UK-based cost studies, cost-of-illness studies and quality of life studies.

The flow diagram describing the economic evaluation search results and study selection process is shown in Appendix 8.14.1.3.

3861 publications were identified through the database search, of which seventy-seven were considered potentially relevant after title and abstract screening. After full text screening, a further forty-eight publications were excluded. The main reason for exclusion after full text screening was “line of therapy” (n=36). In total, twenty-nine publications met the inclusion criteria for the review. These publications reported the results of twenty-three economic evaluations. Two of these evaluations were budget impact analyses and were excluded from the write-up.

The published evaluations were supplemented by data from past NICE technology appraisals, which were identified through an additional ad-hoc search of the NICE website. This search identified two single technology appraisals (TA225 and TA247) and one multiple technology appraisal (TA195) that reported the cost effectiveness of drug therapy in patients who had failed on a previous TNFi. These evaluations were included in the review alongside the published evaluations.

A summary of included economic evaluations is presented in Appendix 8.14.1.3 (published) and Appendix 8.14.1.4 (NICE appraisals).

5.1.2 Description of identified studies

None of the studies identified in the search reported the cost effectiveness of CZP in patients who had failed on a TNFi.

Of the 21 included evaluations, the majority[91] presented a cost-utility analysis of drug therapy, with results reported in terms of the 'cost per QALY gained'. A further eight studies reported the cost effectiveness of drug therapy, with outcomes measured by the 'cost per patient with low disease activity state (LDAS)'. One study reported the results of a cost-consequence analysis, and included costs presented alongside the outcomes of treatment.

In all of the cost-utility analyses, the financial and health outcomes of treatment were estimated using a health economic model. The structure and design of the models varied, and included individual patient-level models and traditional cohort-based models. There was variation in the choice of comparator therapy, the sequence of drugs considered after treatment failure, the patient population, the perspective of the evaluation, and the time horizon over which financial and health outcomes were calculated.

There was also variation in the country of origin for the included studies. This included European countries (16 studies), United States (2), Canada (2) and Mexico (1). Of the sixteen European studies, two reported the cost effectiveness (or cost-utility) of drug therapy in the UK (Emery et al. [abstract], Kielhorn et al.). A more detailed review of these two studies is provided in the following text.

The study by Kielhorn et al. reported the cost-utility of RTX treatment (versus standard of care) in patients with an inadequate response to bDMARDs. The costs and QALYs were evaluated over a lifetime horizon using a microsimulation Markov model. The perspective of the economic analysis was the National Health Service and Personal Social Services. The conclusions of the study were that RTX is a cost effective alternative to standard of care therapy (ADA followed by IFX), with an ICER of less than £15,000 per QALY gained.

Emery et al. assessed the cost effectiveness of different strategies of bDMARD therapy in patients with an insufficient response to at least one prior TNFi. The costs and effectiveness of four different sequences of bDMARDs were compared over a time horizon of two years:

1) ETA followed by ABA followed by ADA
2) ETA followed by RTX followed by ADA
3) ETA followed by ADA followed by ABA
4) ETA followed by ADA followed by IFX

The results suggest that sequences including ABA appear more cost effective than other sequences including RTX or cycled TNFis. This study was available in abstract form only.

In general, the results of the non-UK studies were similar to those reported for the UK. In particular, RTX was found to either dominate (less costly and more effective) or be considered cost effective compared to other bDMARDs. In most studies, there was limited information on whether the studied populations had contraindications or intolerance to either RTX or MTX and it was therefore challenging to relate the studies to the populations in the scope of the appraisal.

A list of excluded studies and the reasons for their exclusion is summarised in Appendix 8.14.1.7.

In general, studies were excluded for the following reasons:

- Disease: If the disease activity (moderate to severe) was not clear
- Intervention: If no specific intervention was present in the study
- Line of therapy: If patients were not previously treated with any TNFi or were not controlled to these drugs.

A quality assessment for each study included in this submission is provided in Appendix 8.14.1.8.

5.2 *De novo analysis*

5.2.1 Patient population

The population of interest to this submission is defined as patients with active moderate to severe RA whose disease has not responded adequately to a TNFi. Moderate to severe disease activity is defined as a DAS28(ESR) score >3.2. In line with the scope of

the appraisal, the cost effectiveness evaluation considers the use of CZP in three populations;

A. Adults previously treated with other DMARDs including at least one TNFi
B. Adults for whom RTX is contraindicated or withdrawn
C. Adults for whom RTX therapy cannot be given because MTX is contraindicated or withdrawn

Due to limited data, the cost effectiveness evaluation does not consider the use of CZP in the 4th population mentioned in the scope, of patients with moderate to severe, active disease despite treatment with bDMARDs.

Baseline characteristics of the modelled population were based on mean estimates from the REALISTIC trial, which was the primary source of data on the efficacy of CZP in the cost effectiveness analysis of CZP in TNFi-IR. There was insufficient data to differentiate the populations based on prior RTX or MTX use, and thus the same baseline characteristics were assumed to apply to all three populations (Table 61). The characteristics of this patient group are in line with the marketing authorisation for CZP in RA.

Table 61: Baseline characteristics of the model population (based on full analysis set)

Characteristic	All patients in the study (CZP and PBO) who are TNFi-IR
Sample**	[redacted]
Mean age, years	[redacted]
Gender, % female	[redacted]
Baseline HAQ score, (0-3)	[redacted]
Baseline Pain score on visual analogue scale, (0-100)	[redacted]
Baseline EQ-5D*	[redacted]
Disease duration, years	[redacted]
At least one prior TNFi One Prior TNFi Two or more prior TNFi	[redacted]

Note: SD= Standard Deviation, * Based on data from PREDICT study, ** based on full analysis set
Source: UCB data on file (RA2015_033_001_04)

5.2.2 Model structure

A cohort-based Markov state transition model was developed to assess the cost effectiveness of CZP in TNFi-IR. The structure of the model is similar to other models in RA, including the cost effectiveness model supporting CZP in DMARD-IR in TA375 and

TA186. As in previous submissions, the model was programmed in Microsoft Excel, and was supported by macros written in visual basic for applications.

The model structure is designed to capture the costs and health effects of the lifetime treatment of TNFi-IR in RA, and comprises a series of health states (including tunnel states) designed to capture the benefits of first and subsequent lines of therapy in this population. These health states include states for response to first therapy (ie. CZP), a series of states used to capture the costs and effects of therapies given after the failure of first treatment (ie. subsequent lines), and a state for death. An illustration of the model structure is provided in Figure 57.

The costs and health effects of treatment are calculated by multiplying the time spent in each state by the costs and health benefits assigned to that state. A half-cycle (or lifetable) adjustment is used when estimating the total costs and health effects of treatment. As per the NICE reference case, the model includes discounting of both costs and effects at a rate of 3.5% per annum.

The health effects of treatment are modelled in terms of the QALY, and are calculated by assigning utility weights to each state in the model. These utility weights include trial-reported EQ-5D utilities (first cycle) and utilities derived using published mapping algorithms (all subsequent cycles).

The health state costs are largely based on treatment assignment, and include the costs of drug acquisition, monitoring, and administration. Hospitalisation costs for RA and related co-morbidities are also included in the analysis. Unlike drug costs, the costs of hospitalisation are modelled based on HAQ score, with increasing HAQ being associated with an increasing cost of hospitalisation. As per the NICE reference case, base case costs and health outcomes are evaluated from the perspective of the NHS and Personal Social Services (PSS).

In addition to NHS and PSS costs, the model also includes the wider societal costs associated with RA. Examples include the financial burden of absenteeism from work and the inability to perform usual daily activities because of disability. In the model, these costs are considered as part of the model's societal perspective and are evaluated in a sensitivity analysis to the NICE base case. As with hospitalisation costs, the wider societal costs of RA are modelled based on HAQ score.

The cycle length of the model varies over the course of the modelled time horizon. The duration of the first cycle is six months and this is based on NICE recommendations on the period for assessing treatment response in RA patients in England and Wales.[92] The second and third cycles are of three month duration, and cover the period between six months and one year. After one–year, all subsequent cycles are of six month duration, which reflects the frequency of monitoring recommended by NICE and the

British society of rheumatology.[51] All model parameters, including costs, health effects, and transition probabilities are adjusted for the time period covered by each model cycle. The maximum time horizon in the model is 45 years.

A summary of the features of the de novo analysis is presented in Table 62.

Table 62: Features of the de novo analysis

Factor	Chosen values	Justification
Time horizon	45 years	The time horizon in the base case analysis is the lifetime of a patient. Analysis of BSRBR data shows an average age of patients starting on TNFis of 55 years.[93] Therefore a timeframe of 45 years captures these patients up to the age of 100 years. Shorter timeframes of 5 and 10 years are considered in the sensitivity analyses.
Were health effects measured in QALYs; if not, what was used?	Yes	
Discount of 3.5% for utilities and costs	3.5%	Sensitivity analyses were performed where discounting was set to 1.5% and 6%
Perspective (NHS/PSS)		The model considers the costs and effects of treatment from the perspective of the NHS and Personal Social Services (PSS), and includes direct medical costs such as hospital care (inpatient and outpatient), primary care and home visits. Sensitivity analyses were conducted using a societal perspective.
PSS, personal social services; QALYs, quality-adjusted life years		

An overview of the clinical pathway in the model is provided in the following section.

5.2.3 Clinical pathway

Patients enter the model after an inadequate response to a TNFi and are assumed to immediately start treatment with a bDMARD (ie. CZP). The initial clinical benefit to treatment is assessed at 6 months. At the end of the initial period, patients are assigned to either one of a series of responder states or to the death state. In the base case, the clinical response states are modelled using EULAR response criteria, and comprise states for “non-responder”, “moderate responder”, and “good responder”.

Following NICE recommendations on stopping rules for biologic treatments, patients assigned to the non-responder state are assumed to discontinue their first therapy, and immediately start treatment on their next therapy in the sequence. Patients assigned to the moderate or good responder states are assumed to continue treatment beyond the responder phase. Responding patients are assumed to benefit from improvements in

utility and HAQ-DI scores. The improvement in utility is assumed to occur over the first 6 weeks of the responder phase.

Throughout all subsequent cycles of the model, patients are at risk of transitioning to the death state or are at risk of discontinuing their current therapy because of lack of efficacy or toxicity. On discontinuation of treatment, patients are assumed to immediately start treatment on the next therapy in their sequence.

For patients who are in the moderate or good responder states, the probabilities of discontinuing therapy were modelled using persistence data from TNFi-IR patients from the BSRBR. In the base case, the same probabilities were applied to CZP and comparator therapies, on the assumption that after initial response, there is no meaningful difference in the toxicity or persistence of bDMARDs. In line with data from BSRBR, the probabilities of treatment discontinuation were assumed to vary with time.

With the cessation of treatment, it is assumed that a patient loses the initial benefits of treatment such that the mean utility decreases in line with the mean utility gains made during the response phase. This loss in utility is partly offset by the health benefits of subsequent therapy.

As utility data are not available for all therapies in the sequence, it was necessary to model the benefits of subsequent treatment in terms of improvement in HAQ score, and to map HAQ scores to utilities via published mapping algorithms (see section 5.4.3). Several studies have shown that self-assessment of pain on a visual analogue scale and HAQ score are independent predictors of utility, and therefore the model simulates both pain and HAQ score over time.[94, 95] In the sensitivity analysis, pain score was modelled by mapping change in HAQ to change in pain from data presented in a previous NICE multiple technology appraisal (TA375). In the model, the loss of efficacy from previous treatment and the gain in efficacy from subsequent treatment are assumed to occur simultaneously at the point of treatment discontinuation. This assumption applies to all therapies in the model.

The probabilities of discontinuation on subsequent therapies were assumed to vary between the first and subsequent six months of treatment. In the first six months, treatment discontinuation was modelled based on EULAR response with only moderate or good responders continuing therapy beyond this period. During all subsequent six monthly cycles, the probability of discontinuation was modelled based on long-term persistence data. Further detail on the data sources used for treatment discontinuation is provided in 5.3.4.

During all cycles of the model, patients may transition to the death state. The probabilities of death are assumed to increase with increasing age and HAQ score. Age and gender-specific mortality probabilities are modelled based on data from the Office

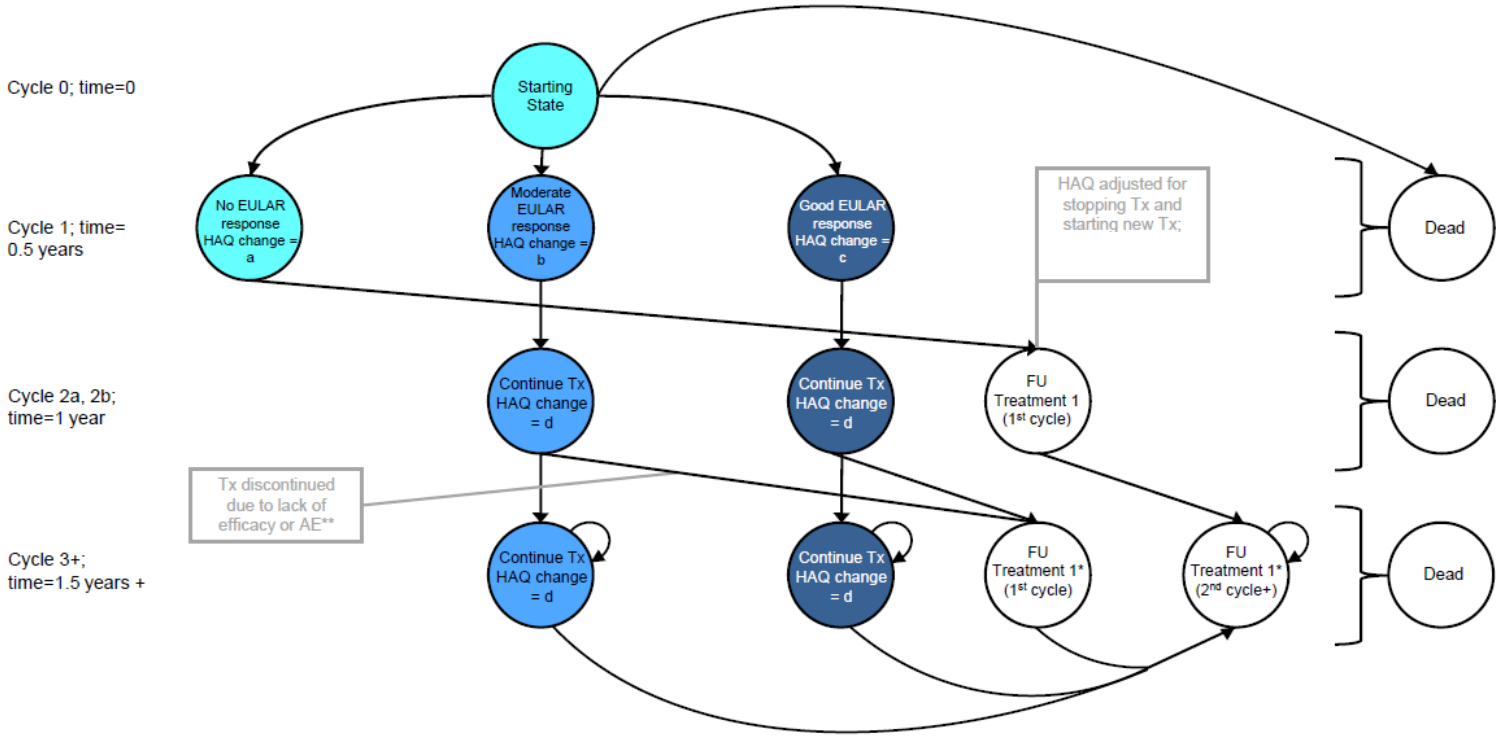
for National Statistics (see section 5.3.6). In the model, all deaths are assumed to occur at the start of each cycle. This includes deaths that occur during the responder phase.

A summary of the health states in the model are shown in Table 63.

Table 63: Health states in the Markov model

Health state/treatment line	Definition	Captures
No response	Failure to respond as per stopping criteria in the UK	Health effects of first treatment in terms of improving the utility of TNFi-IR through achieving EULAR response Stopping criteria for therapy in the UK
Moderate		
Good		
2nd therapy	A sequence of states (tunnel states not reported) representing each subsequent therapy given after failure of first treatment	The costs and health benefits of treatment after failure of first therapy
3rd therapy		
4th therapy		
5th therapy		
6th therapy		
7th therapy		
8th therapy		
Death	-	Effects of disease on life years

Figure 57: illustration of the CZP model structure (excluding transitions to death)



*Follow-up treatment states: duplicated for each follow-up treatment. Patients not responding in first 6 months of follow-up treatment will move to the next treatment in the sequence; **Reason for discontinuation (lack of efficacy) governed by probabilities after leaving treatment health state.
HAQ-DI categories relate to the non-treatment specific costs associated with disability.

5.2.4 Rationale for model structure

A cohort-based Markov-type model was chosen because this type of model is well suited to modelling the prognosis of patients with chronic diseases where events can recur (such as treatment switching), and where the costs and benefits of an intervention accumulate over time. By introducing tunnel states to the model, it is also possible to track the time patients spend in certain states, such as the time on subsequent therapy. These states can be used to incorporate time-varying transition probabilities, and to relax the “memoryless” limitation commonly associated with Markov models. The Markov-type model has been successfully applied to evaluate the cost effectiveness of treatments in RA previously.

Previous Markov models developed for RA have used head-to-head trial data for the proportion of patients in each HAQ-DI state during trial follow-up to model initial response and HAQ-DI change.[96-99] It was not possible or appropriate to use this approach to compare CZP with other biological and cDMARDs, since it requires data that are unlikely to be reported for all comparators, and also it is only possible to describe absolute outcomes. Description of treatment effect in terms of relative risks or odds ratios as required for adjusted indirect comparison would not be possible with this approach.

5.2.5 Intervention technology and comparators

All interventions in the cost effectiveness analysis were used as per their licensed indication and dose administration recommendations. In line with the scope of the appraisal, the cost effectiveness evaluation considers the following comparators to CZP in the TNFi-IR population:

- Population A: RTX in combination with MTX
- Population B: ABA, ADA, ETA, GOL, IFX and TOC each in combination with MTX, for adults for whom RTX is contraindicated or withdrawn
- Population C: ADA monotherapy, ETA monotherapy or TOC monotherapy, for adults for whom RTX therapy cannot be given because MTX is contraindicated or withdrawn

Consistent with previous NICE appraisals and clinical guidelines in England and Wales, the intervention technology and comparators in the economic analysis comprise CZP or comparator(s) followed by a sequence of follow-on therapies. The sequence of follow-

on therapies varies by population, and was selected following consultation with a UK clinical expert. As per the appraisal scope, TOC monotherapy is considered a comparator in population C, although there is currently no guidance on its use in TNFi-IR.

The sequence of follow-on treatments assumed for each population in the cost effectiveness evaluation is presented in Table 64 (Population A), Table 65 (Population B) and Table 66 (Population C).

Table 64: Sequences compared in Population A (patients eligible for RTX plus MTX)

Line of therapy (TNFi-IR)	Sequence with CZP	Comparator sequence
First	CZP + MTX	RTX + MTX
Second	RTX + MTX	TOC + MTX
Third	TOC + MTX	ABA+MTX
Fourth	ABA+MTX	MTX + hydroxychloroquine + sulfasalazine
Fifth	MTX + hydroxychloroquine + sulfasalazine	Non-biologic (Weighted mix of leflunomide, gold, ciclosporin, azathioprine (25% each))
Sixth	Non-biologic (Weighted mix of leflunomide, gold, ciclosporin, azathioprine (25% each))	Palliative care
Seventh	Palliative care	-

Table 65: Sequences compared in Population B (patients for whom RTX is contraindicated or withdrawn)

Line of therapy (TNFi-IR)	Sequence with CZP	Comparator sequences
First	CZP + MTX	Comparator biologic + MTX
Second	MTX + hydroxychloroquine + sulfasalazine	MTX + hydroxychloroquine + sulfasalazine
Third	Leflunomide	Leflunomide
Fourth	Gold injection	Gold injection
Fifth	Ciclosporin	Ciclosporin
Sixth	Azathioprine	Azathioprine
Seventh	Palliative care	Palliative care

Table 66: Sequences compared in Population C (in patients for whom RTX therapy cannot be given because MTX is contraindicated or withdrawn)

Line of therapy (TNFi-IR)	Sequence with CZP	Comparator sequences
First	CZP	Comparator biologic
Second	Leflunomide	Leflunomide
Third	Gold injection	Gold injection
Fourth	Ciclosporin	Ciclosporin
Fifth	Azathioprine	Azathioprine
Sixth	Palliative care	Palliative care

In the model, treatment may be discontinued due to the following events:

- Inadequate response to therapy over the first six months of treatment
- During any subsequent model cycle due to lack of efficacy or adverse events
- Due to death

In line with NICE guidance, an adequate response to treatment is defined by an improvement in DAS28 score of 1.2 points or more, and is defined in terms of EULAR moderate or good response.[92] The assessment of response is made after six months of therapy following recommendations by NICE and the BSR on stopping criteria for biologics in RA.[1] This discontinuation rule is used to identify those patients who benefit the most from treatment, and therefore helps to identify those patients in whom the technology is most cost effective.

EULAR response is used to assess the continuation of existing recommended therapies for TNFi-IR, and its assessment is therefore considered part of routine drug monitoring in this population. As such, there is no additional cost, health, or equity (ie. issues with implementation of the rule in practice) consequences associated with the use of this stopping rule in patients treated with CZP.

Patients that discontinue treatment are assumed to move on to subsequent lines of therapy. The sequence of therapies considered in the cost effectiveness evaluation is detailed in section 5.2.5.

5.3 *Clinical parameters and variables*

The following section contains an overview of the clinical parameters and variables in the model, and includes details on the sources of data used and the approaches to synthesising these data for use in the model. These parameters include:

- Clinical response to first treatment (5.3.1)
- Change in HAQ score associated with first treatment (5.3.2 and 5.3.3)
- Discontinuation of treatment after response to therapy (5.3.4)
- Efficacy and discontinuation of subsequent lines of therapy (5.3.5)
- Mortality associated with RA (5.3.6)

The methods used to estimate the utility improvements associated with response and the mapping of HAQ and pain scores to utility is detailed in 5.4.

5.3.1 Clinical Response to First treatment

5.3.1.1 Summary of clinical response to first treatment

Response to first biologic treatment was modelled based on EULAR response (no response, moderate response and good response) measured at six months. Following approaches adopted in the recent multiple technology appraisal of first-line TNFis in RA, TA375, the probabilities of EULAR response were estimated from a Bayesian NMA of trials identified in the clinical systematic literature review. The statistical model used in the NMA was the conditional binomial likelihood and probit link, which was developed to analyse ordered categorical data, such as EULAR response. Further detail on this model is provided in technical support document 2 from the NICE Decision Support Unit.[82]

In the Excel model, the probabilities of response were estimated using summary statistics on the trial-specific baseline effects, effect size estimates and cut-off statistics extracted from the NMA model. The equations for predicting the probability of response are shown below:

$$P\ (EULAR\ moderate\ ) = \ \Phi(\mu + \beta + \ \mathrm{Z}) - \ \Phi(\mu + \beta)$$

$$P\ (EULAR\ good) = \ 1 - \ \Phi(\mu + \beta + \ \mathrm{Z})$$

$$P\ (no\ response) = \ 1 - \ \mathrm{P}(\mathrm{EULAR\ moderate}) - P(EULAR\ good)$$

Where μ is the trial-specific baseline effects, β is the effect size estimate for treatment versus baseline (ie. PBO), Z is the cut-off statistic and Φ is the inverse of the normal cumulative distribution function. These parameters were varied using a series of univariate normal distributions.

Due to the limited data on the efficacy of biologics in TNFi-IR, it was necessary to make the following assumptions when modelling the efficacy of CZP and comparators in the economic model:

- **Efficacy of CZP and MTX:** In the NMA, the efficacy of CZP was evaluated using data from the REALISTIC study. This was based on a comparison of response probabilities for CZP at Week 28 (six months - open label phase of REALISTIC), versus PBO response probabilities derived by mapping response at Week 12, the end of the PBO controlled phase of REALISTIC, to response at six months via a mapping matrix (further detail is provided in 5.3.1.3)
- **Efficacy of comparator TNFi and MTX:** The systematic literature review identified one randomised controlled trial that evaluated the six month efficacy of a TNFi in TNFi-IR patients; the GO-AFTER study, which reported the efficacy of GOL and DMARD versus PBO and DMARD (subgroup data were reported for MTX treated patients). The review failed to identify studies reporting the efficacy of ADA, ETA, or IFX in TNFi-IR patients. In the absence of data, the efficacy of comparator TNFis were modelled as a treatment class assuming that ADA, ETA, and IFX were of equivalent efficacy to GOL in the TNFi-IR population.
- **Efficacy of biosimiliars to IFX:** Biosimiliars to IFX were assumed to have equivalent efficacy to IFX
- **Efficacy of monotherapy CZP:** The systematic literature review failed to identify any randomised controlled trials reporting the efficacy of biologic monotherapies in TNFi-IR. The only studies to report the efficacy of monotherapy biologics was REALISTIC and HIKARI, which included a subgroup of patients who received monotherapy CZP and monotherapy PBO. The efficacy of monotherapy CZP was modelled using response data from REALISTIC.
- **Efficacy of monotherapy TOC:** The efficacy of monotherapy TOC was modelled by combining the response probabilities for monotherapy CZP, with the effect size estimates for TOC versus CZP derived from the analysis of studies in combination with MTX. This assumes a consistent relative effect for TOC versus PBO, when given as combination or monotherapy in TNFi-IR. The impact of this assumption on the results of the evaluation in population C was assessed as part of the one-way sensitivity analysis (see 5.8).
- **Efficacy of comparator TNFi monotherapies:** As previously, the efficacy of comparator TNFis was modelled as a treatment class assuming that ADA and ETA were of equivalent efficacy to GOL. The efficacies of monotherapy ADA and monotherapy ETA were therefore modelled using the effect size estimates for GOL versus CZP, derived from the analysis of studies in combination with MTX. In sensitivity analyses, ADA and ETA were assumed to be of equivalent efficacy to CZP monotherapy.

A summary of the response parameters used in the model is provided in Table 67. A summary of the data sources used in estimating these parameters is provided in section 5.3.1.2.

Table 67: Summary of clinical data used to predict EULAR response (no response, moderate response, good response) at six months in the economic model

Treatment	Effect size on probit scale (standard error)	Notes/assumptions
Population A		
CZP + MTX	[redacted]	Effect sizes were derived for comparisons of treatment and MTX versus PBO and MTX Probabilities of EULAR response were derived using trial-specific baseline effects [redacted] and cut-off statistic [redacted] for six-month EULAR response in the REALISTIC study. These statistics were combined with the effect size estimates for CZP versus PBO and RTX versus PBO, to provide estimates of EULAR response. Effect sizes, trial-specific baseline effects, and cut-off statistics varied using normal distributions in the probabilistic analysis
RTX + MTX	[redacted]	
Population B		
CZP + MTX	[redacted]	Effect sizes were derived for comparisons of treatment and MTX versus PBO and MTX Due to the lack of data, ADA, ETA, and IFX were assumed to have equivalent efficacy to GOL. Biosimiliars to IFX were assumed to have equivalent efficacy to IFX/GOL Probabilities of EULAR response were derived using trial-specific baseline effects [redacted] and cut-off statistic [redacted] for six-month EULAR response in the REALISTIC study. These statistics were combined with the effect size estimates for CZP versus PBO and comparator versus PBO, to provide estimates of EULAR response. Effect sizes, trial-specific baseline effects, and cut-off statistics varied using normal distributions in the probabilistic analysis
ABA + MTX	[redacted]	
ADA + MTX	[redacted]	
ETA + MTX	[redacted]	
GOL + MTX	[redacted]	
IFX + MTX	[redacted]	
TOC + MTX	[redacted]	
Biosimiliar + MTX	[redacted]	
Population C		
CZP	[redacted]	Effect size were derived for comparisons to CZP, and based on results of the combination therapy NMA (population B)

Treatment	Effect size on probit scale (standard error)	Notes/assumptions
ADA	[redacted]	Due to the lack of data, the efficacies of ADA and ETA were derived using the comparative efficacy of GOL versus PBO, from the combination MTX NMA Probabilities of EULAR response derived using trial-specific baseline effects for CZP monotherapy [redacted] and cut-off statistic [redacted] for REALISTIC study. These statistics were combined with the effect size estimates for TOC versus CZP and TNFi (GOL) versus CZP, to provide estimates of EULAR response. Effect sizes, trial-specific baseline effects, and cut-off statistics varied using normal distributions in the probabilistic analysis
ETA	[redacted]	
TOC	[redacted]	

The predicted probabilities of EULAR response at six months from the Bayesian NMA (multinomial likelihood and probit link model) are presented in Figure 58 (population A), Figure 59 (population B) and Figure 60 (population C).

Figure 58: Population A: Estimated NMA probabilities of EULAR response

Figure 59: Population B: Estimated NMA probabilities of EULAR response

Figure 60: Population C: Estimated NMA probabilities of EULAR response

5.3.1.2 Data used in estimating EULAR response in the Bayesian NMA

A summary of the studies identified in the review, and the reasons for exclusion from the NMA are presented in Table 67. Further detail on the review methodology is provided in Section 4.1.

As noted previously, an NMA was only possible on studies reporting the efficacy of bDMARDs, given in combination with MTX.

Competitor studies

EULAR response probabilities were reported in five studies that evaluated the efficacies of RTX [REFLEX and Combe 2012], ETA [Combe 2012], GOL [GO-AFTER], ABA [ATTAIN] and TOC [RADIATE] at six months. The Combe 2012 study was excluded from the base case NMA because of small sample size (n=10 for RTX and n=10 for ETA), and the associated high risks of a biased effect estimate.

Two studies, GO-AFTER and ATTAIN, reported the efficacy of bDMARD given in combination with cDMARDs (including MTX), and compared to PBO plus DMARDs (including MTX). Subgroup data for the MTX-treated population of GO-AFTER were reported in a post-hoc analysis by Smolen et al, 2014.[100] These data were extracted and used to estimate the efficacy of GOL and MTX in the NMA.

At the time of analysis, there was no subgroup data available for the MTX-treated population of ATTAIN. According to the assessment group report for TA195,[90] the majority of patients in ATTAIN (77.8%) received MTX as their combination therapy. In the absence of subgroup data, the outcomes reported in the ITT population of ATTAIN were incorporated in the NMA.

The review failed to identify RCTs reporting the efficacy of ETA (after excluding Combe 2012), ADA or IFX in TNFi-IR. In the absence of data, these treatments were assumed to have equivalent efficacy to GOL in the economic analysis.

CZP studies

The review identified six studies that evaluated the efficacy of CZP in TNFi-IR:

- REALISTIC
- J-RAPID
- DOSEFLEX
- HIKARI
- PREDICT
- SWITCH

The DOSEFLEX, PREDICT and HIKARI studies were not considered in the NMA because of study design (DOSEFLEX), or lack of a common comparator (HIKARI – no MTX use, PREDICT – no PBO control arm). Further detail on reasons for exclusion is provided in Appendix 8.14.2.1. SWITCH only reported PBO-controlled data until Week 12, in contrast to clinical response at 6 months required by the cost effectiveness analysis.

A subgroup of patients [redacted] in the J-RAPID study had previously received a TNFi. Outcome data relating to this subgroup were extracted. The J-RAPID study was excluded from the base case NMA because of small sample size and the associated risk of a biased effect size estimate. A sensitivity analysis was performed where the J-RAPID study was included in the network of evidence.

The REALISTIC study was a phase IIIb, multicentre trial comprising two phases; a 12-week randomised double-blind PBO-controlled phase followed by an open-label CZP extension. During the initial double-blind phase, patients were randomised to receive either CZP or PBO for 12-weeks. Starting at Week 12, all subjects received treatment with open label CZP for a minimum of 16 weeks (up to Week 28). Randomization was stratified according to concomitant use of MTX, prior TNFi use, and disease duration. Outcome data relating to the TNFi-IR subgroup of REALISTIC were extracted and considered as part of this appraisal.

The design of REALISTIC meant that six month response data (Week 28) were available for CZP but not PBO as the controlled phase of the study concluded at Week 12. This meant that without further adjustment, data collected in the REALISTIC study could not be considered in the NMA, and the study would therefore be omitted from the economic analysis. The omission of both REALISTIC and J-RAPID would exclude all data on the comparative efficacy of CZP versus PBO

in TNFi-IR, and therefore hamper the assessment of CZP cost effectiveness in this population.

To address this data gap, we decided to incorporate REALISTIC in the NMA by estimating the PBO response at six months by mapping response at three months to response at six months. This was achieved by establishing the relationship between response status at three and six months in PBO patients using patient-level data from the RAPID 1 and RAPID 2 studies in DMARD-IR. On the assumption that this relationship can be extended to TNFi-IR, it was then possible to estimate PBO response at six months, using the available three month data (Week 12) and the three versus six month response relationship. A comparison was then made between the estimated PBO response and the actual CZP response at six months. This comparison provided an estimate of the efficacy of CZP versus PBO at the six month assessment. Further detail on this analysis is provided in later sections.

The only studies to report the efficacy of monotherapy biologics was the CZP studies, REALISTIC and HIKARI, which included a subgroup of patients who received monotherapy CZP and monotherapy PBO. In the economic analysis, the probabilities of response for monotherapy bDMARDs were generated by combining the absolute response probabilities for CZP monotherapy with data on the efficacy of TOC versus CZP from the NMA of biologics in combination with MTX. The efficacies of ETA and ADA monotherapy were modelled using the effect size estimates for GOL (versus CZP), on the assumption that these therapies are of equivalent efficacy to GOL (NICE TA225).

Further detail on the NMA methodology and the data used in the analyses is provided in the following subsections.

5.3.1.3 Estimating the comparative efficacy of CZP versus PBO using data from the REALISTIC study

The comparative efficacy of CZP versus PBO at week-28 (or six months) of REALISTIC was derived by comparing response probabilities for CZP at six months, to six-month response probabilities for PBO derived via a mapping analysis.

The baseline characteristics and response probabilities (at Weeks 12 and 28) for the CZP and PBO populations of REALISTIC are presented in Table 68. Only data for the combination therapy group were incorporated in the NMA, due to a lack of comparative data for monotherapy treatments.

Table 68: Baseline characteristics and response to therapy (EULAR and ACR)

Characteristic	CZP + MTX	PBO + MTX	CZP monotherapy	PBO
Data set	Open-label population (treated with CZP in both randomised and open-label phases)	Randomised population	Open-label population (treated with CZP in both randomised and open-label phases)	Randomised population
Baseline characteristics				
Sample size	[redacted]	[redacted]	[redacted]	[redacted]
Age, mean (SD)	[redacted]	[redacted]	[redacted]	[redacted]
Gender, % female	[redacted]	[redacted]	[redacted]	[redacted]
Weight (kg), mean (SD)	[redacted]	[redacted]	[redacted]	[redacted]
BMI (kg/m2)	[redacted]	[redacted]	[redacted]	[redacted]
Concomitant MTX use Yes	[redacted]	[redacted]	[redacted]	[redacted]
Prior TNFi use 1 prior TNF 2 prior TNF >2 prior TNF Unknown	[redacted]	[redacted]	[redacted]	[redacted]
Disease duration, mean (SD)	[redacted]	[redacted]	[redacted]	[redacted]
Patient assessment of pain, mean (SD)	[redacted]	[redacted]	[redacted]	[redacted]
HAQ-DI, mean (SD)	[redacted]	[redacted]	[redacted]	[redacted]
DAS28 score, mean (SD)	[redacted]	[redacted]	[redacted]	[redacted]
Efficacy at three months (Week 12)				
Probability of ACR20	[redacted]	[redacted]	[redacted]	[redacted]
Probability of EULAR moderate or good response	[redacted]	[redacted]	[redacted]	[redacted]
Efficacy at six months (Week 28)				
Probability of ACR20	[redacted]	[redacted]	[redacted]	[redacted]
Probability of EULAR moderate or good	[redacted]	[redacted]	[redacted]	[redacted]

Source: UCB data on file (RA2015_033_001_06, RA2015_033_001_05, RA2015_033_001_01, RA2015_033_001_02, RA2015_029_038_01, RA2015_029_022_01, RA2015_029_004_04, A2015_029_004_06, RA2015_029_004_01)
Note: * randomised population comprises n=207, ** randomised population comprises n=79
Open label extension population

On average, the CZP and MTX group were older [redacted] had longer disease duration [redacted] and had received a greater number of prior TNFis [redacted] than the PBO and MTX group. Despite shorter disease duration, the PBO and MTX group had inferior mean baseline scores for self-assessment of pain [redacted] and HAQ [redacted] when compared with the CZP and MTX

group. In terms of DAS28 score, the main component of EULAR response, there was no meaningful difference in mean baseline score between CZP and PBO groups [redacted] For the purpose of analysing EULAR response, the baseline characteristics of the CZP and PBO groups were considered sufficiently comparable to provide robust estimates of the comparative efficacy of CZP in TNFi-IR.

The CZP monotherapy and PBO monotherapy groups were comparable in terms of mean baseline pain score, HAQ-DI, and DAS28. As in the combination group, patients randomised to PBO monotherapy had shorter mean disease duration and had received fewer prior TNFis than the CZP monotherapy group. These data were not considered in the NMA owing to a lack of evidence on the efficacy of non-CZP biologics in TNFi-IR.

The six month response probabilities for the PBO group was estimated using mapping matrices generated from patient-level data collected in the RAPID 1 and 2 studies (DMARD-IR). A simple transition matrix detailing the numbers and probabilities of transition between responder states over the three and six month assessment visits were generated.

Details on the mapping matrices are provided in Appendix 8.14.2.2.

5.3.1.4 Validation of the comparative efficacy of CZP versus PBO from REALISTIC

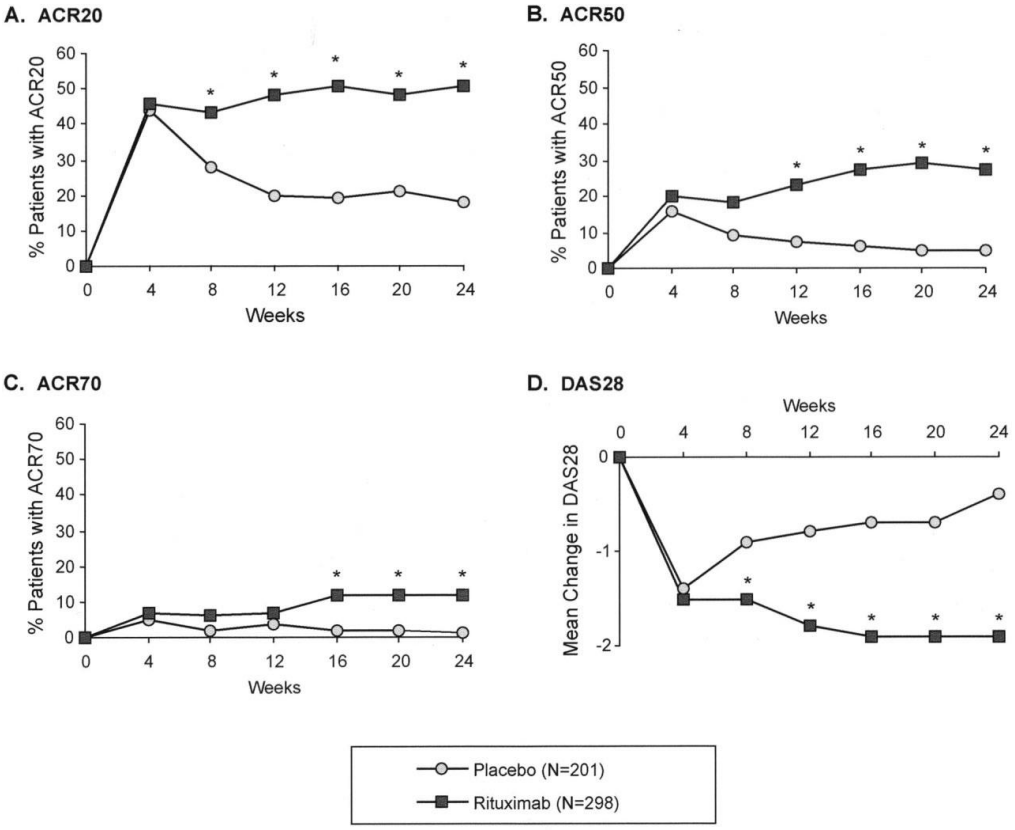
The improvement in comparative efficacy for CZP between three and six months is supported by data from Boers 2010,[101] a meta-analysis comparing the comparative efficacy of experimental treatments in RA (versus control) over the same period.

The main outcome of the Boers study was the pooled ratio of relative risks (experimental versus control) at six versus three months. Ratios of greater than one indicated an improvement in comparative efficacy over this period. Two TNFi-IR studies were included in the meta-analysis (REFLEX and RADIATE). In the TNFi-IR population, the pooled ratios were greater than one in all categories, with values of 1.4 (95% confidence intervals: 1.0 to 2.0) for ACR20, 1.4 (95% confidence intervals: 0.9 to 2.0) for ACR50 and 1.6 (95% confidence intervals: 0.1 to 29.3) for ACR70. These data suggest that continued assessment of response from three to six months in TNFi-IR would yield a numerical improvement in the comparative efficacy of biologic versus PBO. These data therefore support the modelled improvement in the comparative efficacy of CZP versus PBO between the three and six month assessments of REALISTIC.

The modelled decrease in the probabilities of response with PBO at six versus three months is also supported by data from REFLEX, which reports ACR and EULAR response at both three and six month assessments.

As shown in Figure 61, the probabilities of ACR response in the PBO group peaked at week 4 and declined with longer follow-up. The same trend is observed in the mean change since baseline in DAS28 scores (the core component of EULAR response). At six months, the mean change since baseline in DAS28 scores was approximately half the value observed at three months (PBO patients). The probabilities of EULAR moderate or good response at six and three months were 21.0% and 32.5% respectively. The corresponding probabilities of response for RTX were found to increase steadily over the study period in a similar manner to values reported with CZP.

Figure 61: Obtained from REFLEX study - percentages of patients achieving a response according to the American College of Rheumatology 20% improvement criteria (ACR20), 50% improvement criteria (ACR50), and 70% improvement criteria (ACR70), respectively, and D, changes in scores on the Disease Activity Score 28-joint assessment for swelling and tenderness (DAS28) over the 24-week study period in the intent-to-treat population.



The probabilities of response for CZP at Week 28 were compared to the probabilities of response for PBO estimated via the mapping analysis, and then included as part of a Bayesian NMA of response probabilities in TNFi-IR. The results of this NMA are presented in the following section.

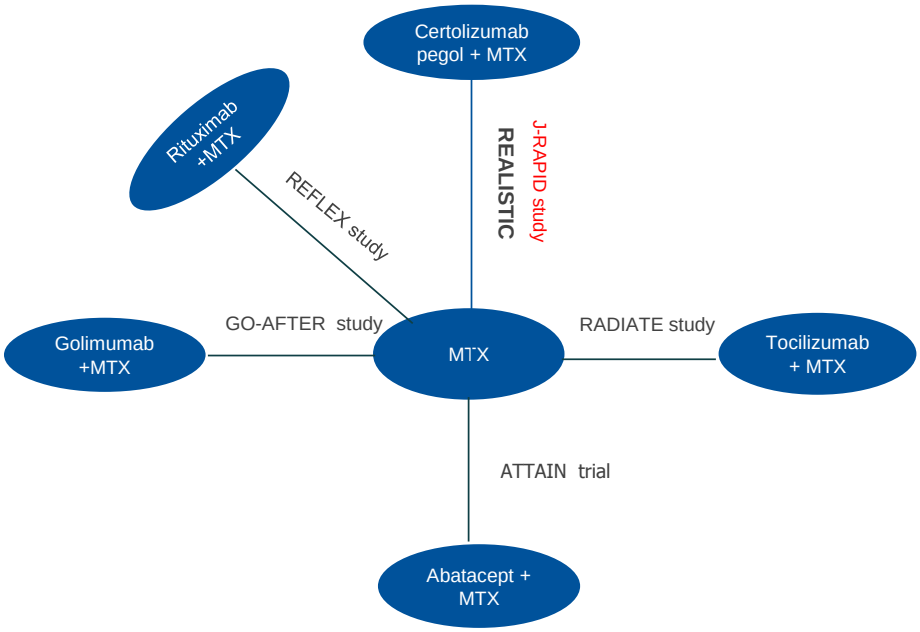
5.3.1.5 Network meta-analysis of EULAR response data

The NMA was performed using the Bayesian statistical program; Openbugs. Due to the limited number of studies in the network (maximum of six), only fixed effect models were considered in the NMA. The results of the NMA were based on 210,000 iterations on three chains after a burn-in of 90,000.

The code for the NMA is provided in Appendix 8.14.2.4, and is based on the code presented in Appendix Six of the technical support document 2.

An illustration of the evidence network for the base case NMA is shown in Figure 62.

Figure 62: Network of evidence considered in the base case and senstivity analysis (including J-RAPID)



A summary of the available data presented in each study is provided in Appendix 8.14.2.3.

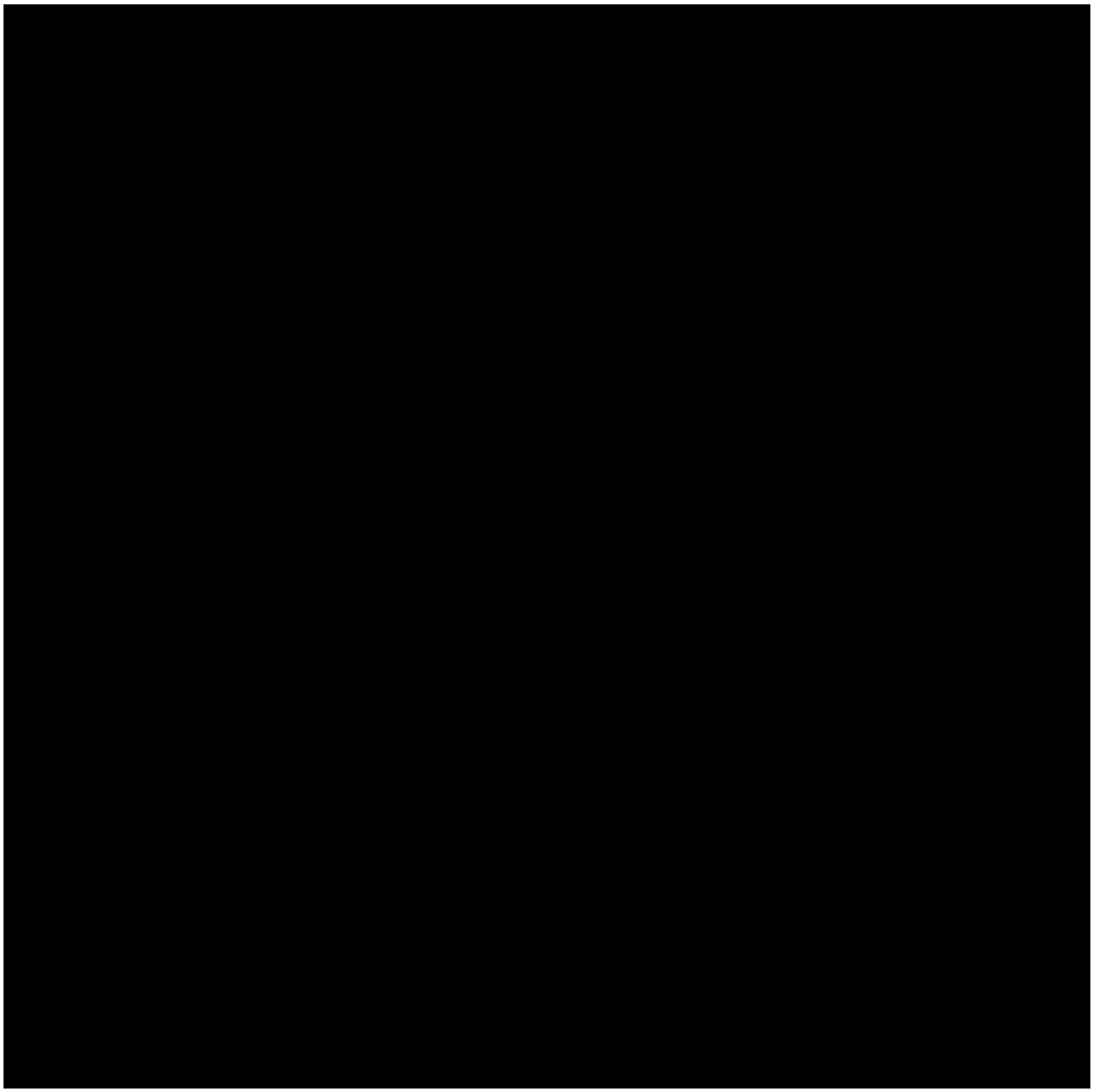
EULAR response data were reported at different cut-offs (moderate, good, or moderate or good) in the six studies considered in the base case NMA. In four of the six studies, EULAR response was reported separately for moderate response, and good response. In two studies, RADIATE and GO-AFTER, EULAR response was reported for a combined group of moderate or good responders. All six studies were incorporated in the NMA on the assumption that the treatment effect is the same regardless of cut-off point. The validity of this assumption was checked by assessing the model's goodness of fit. An acceptable model fit would therefore indicate that this assumption holds.

The results of the base case NMA are presented in Figure 63 and Table 69.

The model fitted the data well, with a total residual deviance of [redacted] in the analysis. As such, the assumption of a common treatment effect across different cut-off points appears valid.

A forest plot showing the effect size estimates on the probit scale for comparisons of bDMARDs versus PBO, and CZP versus alternative biological DMARDs is presented in Figure 63.

Figure 63: Forest plot results of NMA showing treatment effect on the probit scale (EULAR response at six months) comparing biologic versus PBO, and versus CZP



Note: Top, negative value indicates biologic more efficacious than PBO
Bottom, negative value indicates CZP more efficacious than alternative
Effect size estimates reported in Table 69

The NMA results showed that:

- All bDMARDs are associated with a significantly (at the 5% level) increased probability of EULAR response compared to PBO, when given in combination with MTX
- Excluding unlicensed regimens (ie. GOL 100mg and TOC 4mg/kg), CZP + MTX is:

[redacted]

On average, the most efficacious licensed regimens in the NMA were [redacted]

The results of the NMA are in line with the outcomes of the trials included in the analysis.

Table 69: Comparative efficacy of treatment derived from the base case NMA (treatment effect parameters on probit scale), comparing biological versus PBO and MTX, and biologic versus CZP and MTX

Comparison	Mean	L95% CI	U95% CI
Treatment effect on the probit scale comparing biologic + MTX versus PBO + MTX (negative value indicates comparator is more efficacious than PBO)			
RTX + MTX versus PBO + MTX	[redacted]	[redacted]	[redacted]
TOC 8mg + MTX versus PBO + MTX	[redacted]	[redacted]	[redacted]
TOC 4mg + MTX versus PBO + MTX	[redacted]	[redacted]	[redacted]
ABT + MTX versus PBO + MTX	[redacted]	[redacted]	[redacted]
GOL 50 + MTX versus PBO + MTX	[redacted]	[redacted]	[redacted]
GOL 100 + MTX versus PBO + MTX	[redacted]	[redacted]	[redacted]
CZP + MTX versus PBO + MTX	[redacted]	[redacted]	[redacted]
Treatment effect on the probit scale , CZP + MTX versus comparator (negative value indicates CZP is more efficacious than comparator)			
RTX + MTX	[redacted]	[redacted]	[redacted]
TOC 8mg + MTX	[redacted]	[redacted]	[redacted]
TOC 4mg + MTX	[redacted]	[redacted]	[redacted]
ABT + MTX	[redacted]	[redacted]	[redacted]
GOL 50 + MTX	[redacted]	[redacted]	[redacted]
GOL 100 + MTX	[redacted]	[redacted]	[redacted]
PBO parameters			
Trial-specific baseline parameter	[redacted]	[redacted]	[redacted]
Cut-off parameter	[redacted]	[redacted]	[redacted]

Source: NMA analysis

As outlined previously, a sensitivity analysis was performed where the J-RAPID studies were incorporated in the network of evidence. A summary of the results of this analysis are presented in the following paragraphs.

As shown in Appendix 8.14.2.5, the inclusion of J-RAPID had a small positive impact on the predicted efficacy of CZP versus PBO, and CZP versus alternative biologic (including RTX). [redacted]

[redacted]

5.3.2 Change in HAQ and Pain score on response to first therapy

As outlined previously, patients with a moderate or good response to first therapy are assumed to benefit from improvements in utility, HAQ and pain scores. The modelled improvements in utility are derived from EQ-5D data collected in the PREDICT study, as outlined in 5.4. The change in HAQ is modelled based on HAQ scores collected in the REALISTIC study. The change in pain score is predicted from change in HAQ, using data from a previous NICE MTA. Further detail on the parameter estimates are provided in the following paragraphs.

The mean improvement in HAQ score is estimated for each response status through a linear regression model fitted to patient-level data from the REALISTIC study. A regression modelling method was adopted in order that the economic model's probabilistic analyses could incorporate the correlation in HAQ scores across responder groups, and that it was possible to account for differences in baseline characteristics between responder groups (ie. if responding patients have a better baseline HAQ than non-responder patients) when predicting absolute HAQ in the model. Further detail on the methods and fit of the linear regression model is provided below.

The objectives of the regression modelling were to predict the change in HAQ (dependent variable) from baseline to six months, as a function of EULAR response and baseline characteristics (independent variables). Potentially relevant baseline characteristics included demographics (age and gender), disease history (number of prior TNFis and disease duration), baseline disease severity (HAQ and pain scores), and concomitant use of MTX.

A series of bivariate linear regression analyses were performed to assess the association between baseline characteristics and change in HAQ, after adjusting for the effects of response status. Those characteristics considered to be significantly associated with HAQ score in the bivariate analyses (response and baseline characteristic) (p-value <0.05) were considered as part of a multivariate analysis. A stepwise backward routine was then used to optimise the multivariate analysis by removing variables that did not contribute to the predictive validity of the model. The predictive validity of the model was measured via the Akaike information Criterion (AIC). The final regression model was chosen by comparing AIC scores across different potential model specifications, ie. including or excluding baseline variables. The model specification with the lowest AIC score was chosen as the final model.

The mean parameter estimates and p-values for each baseline characteristic considered in the bivariate regression models are provided in Table 70. For brevity, the parameter estimates for response status are not reported.

Table 70: Summary of mean estimates and p-values for selected baseline variables, after adjustment for response status (grey = statistical significant at 5% level)

Model	Parameter	Mean estimate	p-value
Response + Age	Age (years)	[redacted]	[redacted]
Response + gender	Gender (female versus male)	[redacted]	[redacted]
Response + number of prior TNFis	Number of prior TNFis	[redacted]	[redacted]
Response + disease duration	Disease duration (years)	[redacted]	[redacted]
Response + baseline HAQ	Baseline HAQ-DI	[redacted]	[redacted]
Response + Pain	Baseline Pain score on Visual Analogue Scale (0 to 100)	[redacted]	[redacted]
Response + concomitant MTX	Concomitant MTX at baseline	[redacted]	[redacted]

Source: UCB data on file

Baseline HAQ, pain, and concomitant use of MTX were significantly (p-value <0.05) associated with HAQ score, after adjusting for response status. There was no independent association between age, gender, number of prior TNFis, disease duration, and change in HAQ score (p-values > 0.45 after adjusting for response status). Only variables independently associated with HAQ were considered in the multivariate analyses.

In the backward stepwise routine, the model with the lowest AIC value was the full multivariate model comprising all three baseline parameters and response status (AIC=- -429.2065). The stepwise exclusion of concomitant MTX use (AIC=-428.6524), baseline pain (AIC=-426.4389), and baseline HAQ score (AIC=-374.7802) from the full multivariate led to inferior AIC scores compared to the full model. Hence, the full multivariate analysis was chosen as the final model for the economic analysis.

A summary of the final linear regression parameters for predicting change since baseline HAQ at Week 28 in REALISTIC is presented in Table 71.

Table 71: Final linear regression parameters for change since baseline in HAQ-DI score at Week 28 in REALISTIC

Parameter	Mean estimate	Standard error	95% lower CI	95% Upper CI	P-value
Intercept	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Moderate responder*	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Good responder*	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Parameter	Mean estimate	Standard error	95% lower CI	95% Upper CI	P-value
Baseline HAQ score	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Baseline Pain score on Visual Analogue Scale (0 to 100)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Concomitant use of MTX versus other DMARDs or monotherapy	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Information:

[redacted]

[redacted]

*: non-responder is the referent category in the regression analysis, and has a fixed coefficient of zero in the economic analysis

A plot showing the distribution of residuals from the final model is shown in Appendix 8.14.2.6, along with further information.

The fit of the linear regression model in predicting the digitised points on the HAQ versus pain chart is shown in Figure 64.

Figure 64: fitted function of HAQ against pain based on data from the National Data Bank

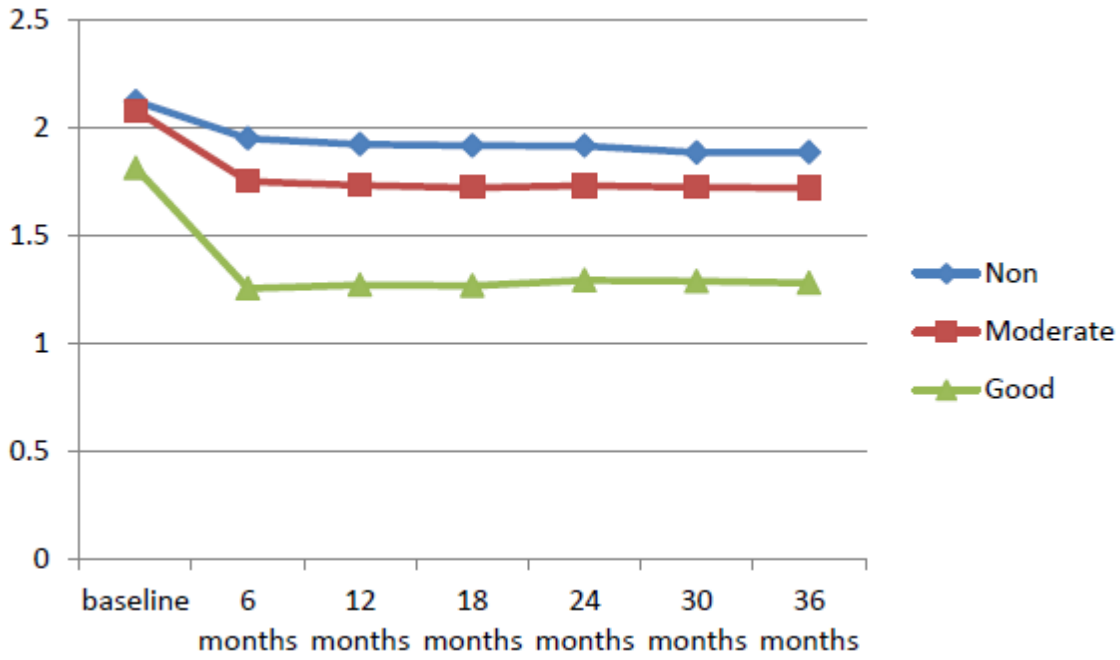
[redacted]

The intercept and slope parameters for the linear model were [redacted] and [redacted] respectively. The slope parameter for the model provides an estimate of the change in pain score associated with a unit change in HAQ. This parameter estimate was used to predict pain from HAQ over time in the model.

5.3.3 Change in HAQ and Pain score after response to first therapy

In line with previous NICE appraisals, it was assumed that the mean HAQ, utility and pain scores of patients who remain on biologic treatment would remain constant up to the point of treatment discontinuation. This is supported by data from the BSRBR and reported in the assessment group report for TA375, which shows stable mean HAQ scores up to three years after initiation of biologic therapy (Figure 65).

Figure 65: Mean HAQ by EULAR response category for those receiving bDMARDs in the BSRBR and reported in figure 110 of the assessment group report for NICE TA375



Limited data are available on the long-term HAQ score of TNFi-IR receiving biologic therapies. Efficacy data from the extension phases of the ATTAIN (ABA and DMARD),[102] and REFLEX studies (RTX and MTX)[103] suggest that response to biologic therapy can be maintained from six months to up to five years post-treatment in TNFi-IR.

5.3.4 Discontinuation of treatment after response to first therapy

The probability of treatment discontinuation following response to first therapy was modelled based on discontinuation data from TNFi-IR patients registered to the BSRBR. These data were originally reported by the BSRBR in their submission to

TA195.[90] These data were re-analysed to provide input parameters to the assessment group's model. The results of the re-analysis are reported in the technical appendix to TA195.[90]

As outlined in TA195, the probability of treatment discontinuation for second-line TNFis were modelled by fitting a Weibull distribution (time in years) to a digitised copy of the Kaplan-Meier plots of discontinuation of second-line TNF (for any reasons).The technical appendix to TA195 contains limited detail on the fit of the model to the digitised data from the BSRBR. The only indication of model fit is provided through a simple graphic showing the fit of the Weibull model to selected data points from the BSRBR. Based on visual inspection of this graphic, it appears that the model provides a robust fit to the data.

The technical appendix to TA195 details the mean estimates and standard errors for the shape and scale of the Weibull distribution, but does not report the variance-covariance matrix. As such, it was not possible to sample these parameters from a multivariate distribution, as recommended in the literature. In the absence of the variance-covariance matrix, the Weibull parameters were sampled via univariate normal distributions and therefore assuming no correlation between parameters.

The mean estimates and standard errors for the shape and scale of the Weibull distribution are shown in Table 72.

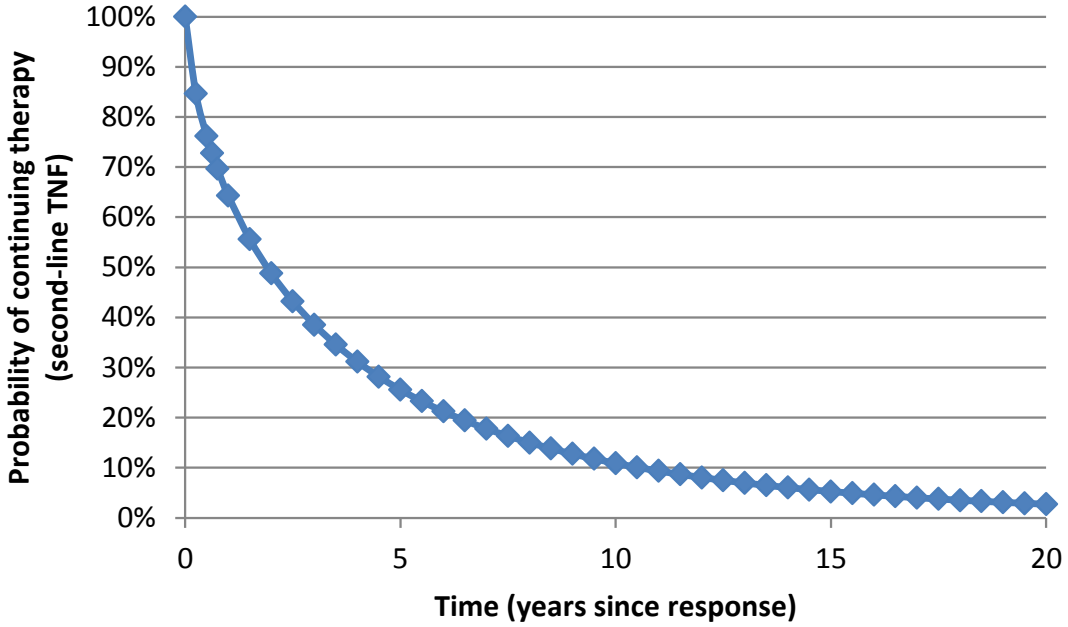
Table 72: Weibull parameters* for the continuation of biologic therapy after response

Parameter	Estimate	Standard error
Scale	0.441555	0.00958300
Shape	0.7008	0.033681
Equation: S(t) =exp (- scale x (t)^ shape)		

Source: Malottki et al (2011)[90]

A graphical illustration of the predicted survival curve for continuation of second-line TNF is shown in Figure 66. The estimated mean duration of second-line TNF was approximately 4.3 years.

Figure 66: Weibull function for probability of treatment discontinuation



Source: UCB model

In the model, the cycle-specific transition probabilities for therapy discontinuation were estimated from the cumulative survival function via the following equation:

$$S(t) = 1- (S(t+\alpha)/S(t))$$

Where t is the time point at the start of the cycle and α is the cycle length.

The transition probabilities for therapy discontinuation were assumed to apply from the second cycle of the model, as treatment discontinuation during the first cycle was captured through EULAR response. The transition probabilities from the six month point of the BSRBR curve were applied at the six month point of the model.

As noted previously, the same probability of treatment discontinuation after clinical response was applied to CZP and all comparators in the evaluation. This was based on the assumption that after clinical response, there is no meaningful difference in the toxicity or persistence across different bDMARDs.

Following consultation with a clinical expert, it was noted that RTX may be associated with a longer duration of therapy than other biologics given that it is administered at irregular intervals. Therefore, a sensitivity analysis was performed assuming lower rates of discontinuation in patients receiving RTX and other non-TNFi biologics (ABA and TOC), versus TNFis.

The lower rate of discontinuation for non-TNFi biologics are modelled using data from Du Pan et al, a study assessing drug retention rates between TNF and non-TNFis in 1485 TNFi-IR registered to the Swiss RA (SCQM-RA) registry. Du Pan et al report a 50% reduction in drug discontinuation risk in favour of non-TNFis, with an

adjusted hazard ratio of 0.50 (95% confidence interval of 0.41 to 0.62). In the model sensitivity analysis, the scale parameter for the discontinuation survival curves for non-TNFis were proportionally reduced by 50% to account for a potentially longer duration of treatment for these therapies. Sensitivity analysis was also performed to test the assumption shorter duration of non-TNFis compared to TNFis, based on findings from Ramiro 2015.[104] Further detail on the sensitivity analysis is provided in Section 5.8.

5.3.5 Efficacy and discontinuation of subsequent therapies

The efficacy and discontinuation of treatment of subsequent treatments were modelled in terms of:

- Change in HAQ score over the first and subsequent six months of treatment
- Probability of treatment discontinuation over the first and subsequent six months of treatment

The change in HAQ score for subsequent therapy was used to predict improvements in utility whilst on therapy. The benefits and associated costs of subsequent therapy were assumed to accrue for the duration patients remained on treatment, which was modelled based on the probabilities of treatment discontinuation.

Limited data were available on the efficacy of therapies after failure on two or more TNFis.[105] In the absence of data, the efficacy of subsequent biologics and cDMARDs were modelled using data from the RADIATE study, where approximately 50% of the enrolled population had received two or more TNFis prior to baseline.[91] Further detail on the modelling of these parameters is provided below.

5.3.5.1 Change in HAQ score for subsequent therapies

The change in HAQ score over the first six months of subsequent biologic therapy was assumed the same for all biologics, and modelled based on efficacy data from the TOC 8mg/kg arm of the RADIATE study (-0.39 mean change in HAQ over six months, standard error=0.04). For all cDMARDs, the change in HAQ score over the first six months was modelled based on data from the PBO and MTX arm of RADIATE (-0.05 mean change in HAQ over six months, standard error=0.01). Palliative care was assumed to be associated with no improvement in HAQ during the first six months.

The change in HAQ score after the first six months of therapy was assumed to be zero for all subsequent bDMARDs. For cDMARDs and palliative care, it was assumed that HAQ scores increased at a rate of 0.045, and 0.06 per annum based on data used in previous NICE appraisals. The maximum mean HAQ score for the population is user definable in the model. In the base case, the maximum mean HAQ is set to 2.76 based on the characteristics of a subpopulation of the US National

Data Bank for Rheumatic Diseases, with long-established RA (mean disease duration of 31-years)[106] The ceiling score for the HAQ is 3.0.

5.3.5.2 Probability of treatment discontinuation for subsequent therapies

The probability of treatment discontinuation on subsequent therapies was modelled using data from several sources.

For the first six months of therapy, discontinuation was based on an inadequate response to therapy using the EULAR criteria.

- For subsequent biologic therapies, the probability of inadequate response at six months was modelled using the treatment effect parameters in the NMA, applied to the trial-specific baseline effects from the RADIATE study (ie. PBO and MTX baseline). The resulting probabilities of response were lower than those considered for first therapy (and derived using trial-specific baseline effects from REALISTIC), but were consistent in terms of predicting higher response for TOC and RTX versus ABA
- For subsequent cDMARDs, the probability of inadequate response was modelled using response data from the PBO and MTX arm of RADIATE (83.7% with inadequate response at six months). The probability of inadequate response for the case mix of non-biologic therapies was derived from 1-year continuation data for MTX that was reported in Edwards et al.[107] These data were converted to six monthly transition probabilities.

For all subsequent six monthly cycles of therapy, treatment discontinuation was modelled based on discontinuation data from the BSRBR (biologics) and from data in Edwards et al (cDMARDs).

For biologic therapies, the probability of discontinuing therapy during each subsequent six monthly cycle in the model was calculated using the BSRBR discontinuation curve reported earlier in the submission document. In the model, the probability of discontinuation is held constant over time and modelled based on the discontinuation rates between six months and one-year in the BSRBR, using the following equation:

$$1-\exp(-scale*(1^{shape}))/\exp(-scale*(0.5^{shape}))$$

For cDMARDs, the probability of discontinuing treatment was estimated from the cumulative probability of treatment persistence at 1-year and 5-years in Edwards et al, and assuming a constant transition probability over this period. To derive the constant transition probabilities, the survival data reported in Edwards et al were converted to cumulative hazard rates using the following equation:

$$H(t) = - \text{Log}(S(t))$$

The six-monthly transition probability for treatment discontinuation was then derived using the following equations:

$$H[0.5] = [H(5) – H(1)] / [4 \text{ years} \times 2 \text{ six month cycles per year}]$$

$$TP = 1\text{-}\exp(-h[0.5])$$

Where TP is the transition probability and h is the instantaneous hazard rate

Cumulative survival probabilities were available for MTX, gold injection, ciclosporin, and azathioprine (Table 73) at 1-year and 5-years post initiation of therapy. Only 1-year survival probabilities were available for leflunomide (84.6% at 1-year). Edwards et al did not report discontinuation for specific combination therapies, such as MTX, hydroxychloroquine, and sulfasalazine, or for the case mix of non-biologic therapies. In the base case, the probability of discontinuation for leflunomide, combination therapy and the case mix of non-biologics was modelled using discontinuation rates for MTX.

A summary of the derived 6-month probabilities for treatment discontinuation are presented in Table 73.

Table 73: 1-year and 5-year survival probabilities for cDMARDs reported in Edwards et al, and estimated cumulative hazard functions, and 6-monthly probabilities and hazard rates

Treatment	Percentage of DMARD users remaining on treatment as reported in Edwards et al		Derived 6-monthly probability/rates**
	At 1-year	At 5-years	
MTX* Probability	78.00%	57.10%	3.8%
Cumulative hazard (derived)	0.248	0.560	0.039
Gold injection Probability	45.90%	17.60%	11.3%
Cumulative hazard (derived)	0.779	1.737	0.120
Ciclosporin Probability	62.00%	34.20%	7.2%
Cumulative hazard (derived)	0.478	1.073	0.074
Azathioprine Probability	56.90%	34.80%	6.0%
Cumulative hazard (derived)	0.564	1.056	0.061

A summary of the efficacy parameters for subsequent therapies in the model are reported in Table 74.

Table 74: Efficacy of subsequent therapies given after initial treatment for TNFi-IR in the economic analysis

Subsequent treatment given after initial treatment	Short-term change in HAQ (six months)	Long-term change in HAQ (six months)	Probability of discontinuation in first 6-months	Probability of discontinuation in all subsequent six monthly cycles
RTX + MTX	-0.39	0	46.6%	15.6%
TOC + MTX	-0.39	0	34.5%	15.6%
ABA + MTX	-0.39	0	61.2%	15.6%
MTX + Hydro + Sulfasalazine	-0.05	+0.02	83.7%	3.8%
Leflunomide	-0.05	+0.02	83.7%	3.8%
Gold injection	-0.05	+0.02	83.7%	11.3%
Ciclosporin	-0.05	+0.02	83.7%	7.2%
Azathioprine	-0.05	+0.02	83.7%	6.0%
Non-biologic therapy (sequence of therapies)	-0.05	+0.02	18.9%	3.8%
Palliative care	0	+0.03	0%	0%

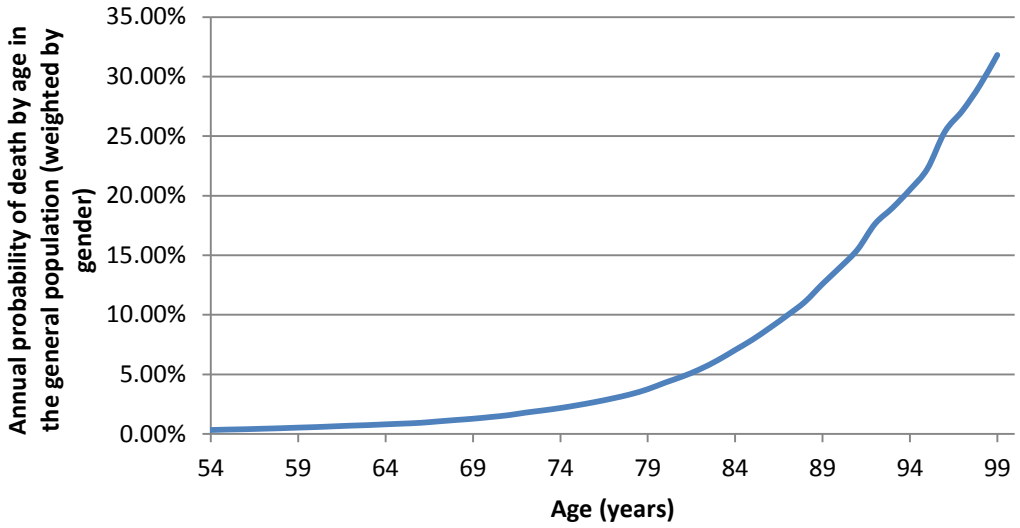
5.3.6 Mortality associated with RA

In the model, the probabilities of death are assumed to increase with increasing age and disability status (in terms of HAQ score).

The relationship between age and mortality was modelled using general population all-cause mortality statistics published by the Office for National Statistics. Age and gender-specific annual probabilities of death were extracted from the latest interim lifetables.[108] A gender-averaged mortality probability was generated based on the proportion of male versus females in the modelled cohort.

A graph showing the gender-averaged mortality probabilities by age is presented in Figure 67.

Figure 67: Plot showing the six monthly probability of death by age in the general population, weighted by the gender distribution in the modelled cohort



The excess mortality associated with RA and its comorbidities were modelled based on HAQ score, with increasing HAQ being associated with an increased risk of mortality. Several studies have shown that the HAQ is the most powerful predictor of mortality in patients with RA,[109, 110] and this approach has been adopted in previous NICE appraisals in RA.

The relationship between HAQ and mortality was modelled using data from Norton et al, a study of the trajectories of functional limitation in early RA and their association with mortality.[111] The study population in Norton et al comprised patients recruited from nine hospitals in different regions of England between 1986 and 1998.[111] Norton et al report that HAQ score assessed at 1 year was a significant predictor (hazard ratio=1.43 [95% confidence interval 1.17 to 1.75]) of mortality after adjustment for differences in clinical and demographic factors.[111]

In the model, the hazard ratio for mortality by unit change in HAQ is used to proportionally increase the hazard rate of death in line with the changing HAQ score of the population. For patients with a HAQ score of 3.0, the rate of mortality was approximately 3-times greater than in the general population (matched for age and gender). The relationship between HAQ and mortality multiplier is shown in Table 75.

Table 75: Modelled relationship between HAQ and mortality, presented in terms of mortality rate multipliers

HAQ	Modelled HAQ multiplier for rate of death by HAQ state
0.0	1.000
0.5	1.196
1.0	1.430

HAQ	Modelled HAQ multiplier for rate of death by HAQ state
1.5	1.710
2.0	2.045
2.5	2.445
3.0	2.924

5.4 *Measurement and valuation of health effects*

Health-related quality-of-life evidence used in the economic model was split into two elements:

- First 6 months: Relationship between initial response from therapy and EQ-5D
- Beyond 6 months: Mapping from HAQ score to EQ-5D

The baseline utility and the function used to map utility to response were derived from a statistical analysis of EQ-5D data from the PREDICT study.

Mapping algorithms were identified from a recent study by Pennington,[112] where different types of mapping methods were introduced and assessed. A summary of the different utility algorithms identified is presented in Section 2.4.3

5.4.1 Health-related quality-of-life data from clinical trials

The EQ-5D questionnaire (five-domains, three-levels) was administered at baseline, at Weeks 6, 12, 24, and at week 52 or withdrawal of the PREDICT study.

In line with the NICE reference case, the EQ-5D domain responses were mapped to utilities using the UK social tariff.[113] Missing assessments were imputed using the last observation carried forward approach. This included missing assessments because of withdrawal due to non-response at Week 12 of the study.

Data were extracted for the subgroup of patients with prior use of TNFis. Completion rates for the EQ-5D were generally good across the study follow-up. Of the 409 patients with past TNF use in PREDICT, utilities were available from [redacted] patients at baseline, and [redacted] patients at Week 24.

A graph showing the mean EQ-5D utility by study visit is presented in Figure 68.

Figure 68: Graph showing the mean EQ-5D utility by visit in PREDICT

Note: TNFi-IR population, imputation using LOCF

The mean EQ-5D utility at baseline, and Weeks 6, 12, 24, and 52 (end of study) are reported in Table 76.

Table 76: summary of mean EQ-5D utility, 95% confidence interval and number of EQ-5D utility measures by study visit in PREDICT

Study time point	Mean EQ-5D utility	95% confidence interval (upper to lower)	Number of EQ-5D questionnaires
Baseline			
Week 6			
Week 12			
Week 24			
End of study / Week 52			

Source: UCB data on file (RA2015_029_059)

As shown in Figure 68, there was an improvement in the mean EQ-5D from baseline in patients treated with CZP, which occurred rapidly after 6 weeks of therapy. The mean utility reported at week 6 was then maintained up to one-year after initiation of treatment.

5.4.2 Mapping

Not applicable as EQ-5D data were collected in the PREDICT study. Further detail on the mapping of HAQ to EQ-5D is provided in section 5.4.3.

5.4.3 Health-related quality-of-life studies

Relevant health related quality of life data were obtained from a review of mapping algorithms in RA that was reported by Pennington and Davis.[112]

The objectives of the review were to identify mapping algorithms that could be used to convert HAQ scores into EQ-5D. The search was intended to identify algorithms that included HAQ score, age, sex and pain measured on a visual analogue scale. Further detail on the search methods is provided in the full publication.[112]

Pennington and Davis identified 24 utility mapping algorithms that could be used to convert HAQ scores into the EQ-5D.[112] A summary of these studies is provided in Appendix 8.14.2.7. For each algorithm, a crude estimate of the change in EQ-5D per unit change in HAQ was derived based on the covariates reported for HAQ score (and excluding non-linear terms).

Eleven of the 24 algorithms reported the regression mapping of HAQ to EQ-5D utility without the consideration of other covariates, such as age or pain. Of these studies, nine were based on linear mapping (EQ-5D = intercept + HAQ) and two were based on non-linear mapping (EQ-5D = intercept + HAQ + HAQ^2). The crude mean change in EQ-5D per unit change in HAQ (excluding non-linear terms) for these models ranged from -0.11 to -0.327.

Eight of the 24 mapping algorithms reported absolute EQ-5D utilities by HAQ categories. All studies reported that increasing HAQ was associated with a reduction in EQ-5D utility. The remaining four algorithms were retrieved from a study by Hernandez Alava et al,[95] all of which considered the mapping of HAQ and pain to EQ-5D utility. The crude mean change in EQ-5D per unit change in HAQ for these models ranged from -0.062 to -0.165 (excluding non-linear terms and the effects of pain).

5.4.4 Adverse reactions

The model did not consider the effect of adverse reactions on health utility on the basis that there is no meaningful difference in the toxicity or risks of adverse events between alternative bDMARDs.

5.4.5 Health-related quality-of-life data used in cost effectiveness analysis

The relationship between clinical outcomes and health utility was modelled in several stages as outlined in the following text:

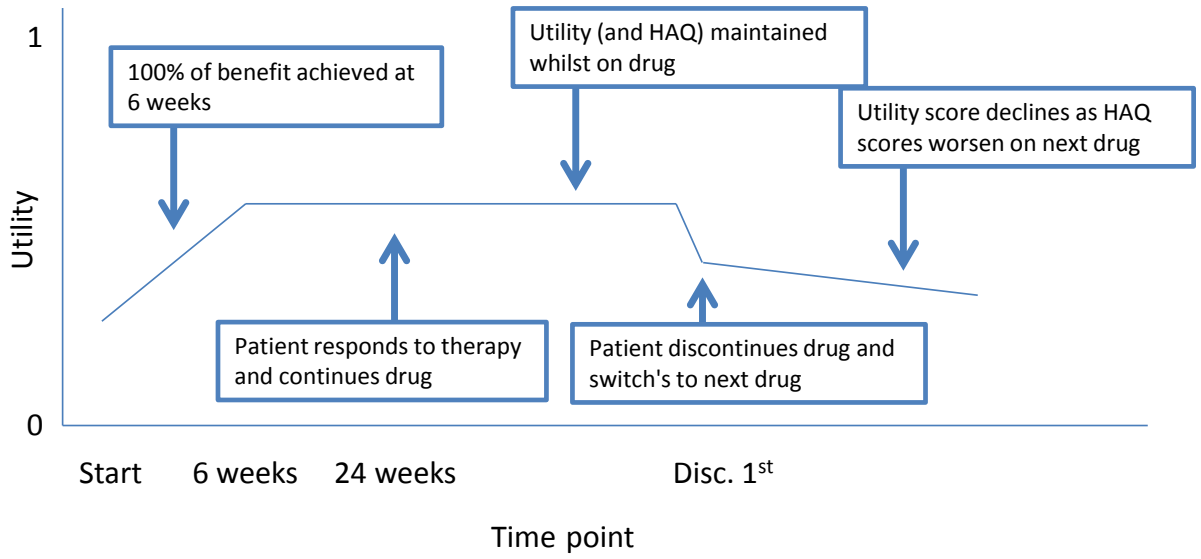
- **Estimates of utility improvements on initial response on first-line treatment.** On entry into the model, the patient population is assigned a mean pre-treatment utility score of 0.4012, derived from the EQ-5D data collected in

the PREDICT trial. Over the first 6 months of treatment, patients are then assigned an average change in the derived EQ-5D utilities which is dependent on response category. The magnitude of the change in EQ-5D utilities was estimated from a regression analysis of patient-level data from PREDICT, and was assumed to be the same for CZP and all comparators. The base case analysis also assumed that 100% of the change over the first six months is achieved by week six, to reflect the rapid response to treatment.

- **Rebound.** Patients discontinuing treatment are assigned a decrease in utility equal to that applied for the initial response to treatment and, immediately, an increase in utility resulting from response to a subsequent treatment. Thus the model does not favour interventions with low discontinuation, since the benefit of initial treatment is replaced by a benefit of follow-up treatments. This assumption is based on previous published RA models.[114] .
- **Utility estimates on subsequent therapies.** The utilities assigned to patients receiving subsequent treatments are modelled based on HAQ (and pain in the sensitivity analysis) that is mapped to EQ-5D utility via a mapping algorithm.

A graphical illustration of the modelling of EQ-5D utilities in the economic model is provided in Figure 69.

Figure 69: illustration of change utility score over time point in the analysis



Note: Disc. 1st refers to discontinuation of 1st drug therapy

A summary of the EQ-5D utilities values used in the cost effectiveness model is provided in Table 77.

Table 77: Summary of utility values for cost effectiveness analysis applied to CZP and comparator sequences

Health state / line of treatment	Utility value: mean (standard error)	Reference in submission	Justification
Baseline utility	[redacted]	5.4.1	Baseline utility from PREDICT study
Non-responder to first therapy	[redacted]	5.4.5	Modelled based on a regression analysis to predict change since baseline EQ-5D utility based on response status, and mean baseline EQ-5D. EQ-5D data were available from a subpopulation of TNFi-IR enrolled to PREDICT
Moderate-responder to first therapy	[redacted]		
Good-responder to first therapy	[redacted]		
Subsequent therapy with bDMARD (ie. RTX, ABA, TOC)	[redacted]	5.3.5 and 5.4.5	As utility data are not available for all therapies in the sequence, it was necessary to model the benefits of subsequent treatment in terms of improvement in HAQ score, and to map HAQ scores to utilities via a published mapping algorithm (Brennan et al). The predicted utilities for subsequent therapy varied from [redacted] for subsequent biologics to a minimum of [redacted] in patients with a maximum HAQ of [redacted].
Subsequent therapy with cDMARD (ie. MTX)	[redacted]		
Subsequent therapy with Palliative care	[redacted]		

The derivation of EQ-5D utilities for stage of the model calculation is presented in the following sections.

Estimates of utility improvements on initial response on first-line treatment

The mean improvement in EQ-5D score is estimated for each response status through a linear regression model fitted to patient-level data from the PREDICT study. A regression modelling method was adopted in order that the model's probabilistic analyses could incorporate the correlation in EQ-5D scores across responder groups, and that it was possible to account for differences in baseline characteristics between responder groups (ie. if responding patients have a better baseline EQ-5D utility than non-responder patients) when predicting absolute EQ-5D utility scores in the model. Further detail on the methods and results of the linear regression modelling of HAQ scores is provided below.

The objectives of the regression modelling were to predict the change in EQ-5D scores (dependent variable) between baseline and six months associated with each response category of the EULAR, with further adjustment for other relevant baseline characteristics that predict EQ-5D (independent variables). Potentially relevant

baseline characteristics included age, gender, number of prior TNFis, disease duration, and baseline EQ-5D utility scores.

A final regression model was developed by conducting a series of linear regression analyses to assess the independent association between each baseline characteristic and change since baseline EQ-5D utility, after adjusting for the effects of response status. Those variables considered to be independently associated with EQ-5D utility in the univariate analyses (p-value <0.05) were considered as part of a multivariate analysis. A stepwise backward routine was then used to optimise the multivariate analysis by removing variables that did not contribute to the predictive validity of the model. The predictive validity of the model was measured via the AIC. The final regression model was chosen by comparing AIC scores across different potential model specifications, ie. including or excluding baseline variables. The model specification with the lowest AIC score was chosen as the final model.

A summary of the parameter estimates and p-values for each of the univariate analyses is provided in Table 70. For brevity, the corresponding parameter estimates for response status are not reported.

Table 78: Summary of mean estimates and p-values for selected baseline variables, after adjustment for response status (grey = statistical significant at 5% level)

Additional parameter to response	Mean estimate	p-value
Age (years)	[redacted]	[redacted]
Gender (female versus male)	[redacted]	[redacted]
Number of prior TNFis	[redacted]	[redacted]
Disease duration (years)	[redacted]	[redacted]
Baseline EQ-5D utility	[redacted]	[redacted]

Source: UCB data on file

Baseline EQ-5D was the only variable significantly (p-value <0.05) associated with EQ-5D utility, after adjusting for response status. There was no independent association between the characteristics of age, gender, number of prior TNFis, disease duration, and EQ-5D utility score (p-values > 0.45 after adjusting for response status). [redacted] Hence, the full model with baseline EQ-5D included was chosen as the final model for the economic analysis.

A summary of the final linear regression parameters for predicting change since baseline in EQ-5D at Week 24 in PREDICT is presented in Table 71.

Table 79: Final linear regression parameters for change since baseline in EQ-5D at Week 24 in PREDICT

Parameter	Mean estimate	Standard error	95% lower CI	95% Upper CI	P-value
Intercept	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Moderate responder*	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Good responder*	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Baseline EQ-5D utility score	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Information:

Number of observations in data: 398
Number of observations used in analysis: 376

Root mean squared error: 0.22100
AIC: -1131.2199

*: non-responder is the referent category in the regression analysis, and has a fixed coefficient of zero in the economic analysis

Source: UCB data on file

A plot showing the distribution of residuals from the final model, and the observed and predicted change since baseline in EQ-5D utility weights, are shown in Appendix 8.14.2.8, along with further information.

Rebound

Patients discontinuing treatment are assigned a decrease in utility equal to 0.117 for non-responders, 0.262 for moderate responders and 0.367 for good responders and, immediately, an increase in utility resulting from response to a subsequent treatment. The increase in utility from subsequent treatment is detailed in the following section.

Utility estimates on subsequent therapies

The EQ-5D utilities assigned to subsequent therapies are estimated based on HAQ scores (and pain in the sensitivity analysis) that are mapped to EQ-5D utility via a mapping algorithm. To ensure consistency between baseline EQ-5D utility, and the utilities assigned to subsequent therapies, we mapped the efficacy of subsequent therapies in terms of improving HAQ score to an estimate of improvement in EQ-5D utilities.

In the base case, HAQ scores are mapped to EQ-5D utilities using the mapping algorithm from Brennan et al (not reported in Pennington and Davis), which reported a mean change in EQ-5D per unit change in HAQ of -0.2012. This data source was used in previous cost effectiveness assessments of CZP in NICE TA375.

A sensitivity analysis was performed using the regression coefficients from the finite mixture models presented by Hernandez Alava et al,[115] and which explore the relationship between EQ-5D, HAQ and pain. The full equations for Hernandez Alava

et al were not implemented in the cohort model to avoid over complicating the model programming. Instead, the coefficients for HAQ (linear term) and pain for each of the four classes of Hernandez Alava et al were programmed in the model with the option of selecting any one of the classes, as required. All other parameters, including non-linear terms for HAQ were excluded from the model. A series of sensitivity analyses were then performed using the covariates from each of the regression classes, to provide a range of potential outcomes based on Hernandez Alava et al.

5.5 *Cost and healthcare resource use identification, measurement and valuation*

5.5.1 Resource identification, measurement and valuation studies

5.5.1.1 Identification of cost and resource use studies

Search strategy

Assessment reports from previous NICE HTAs were searched for relevant studies to inform costs and resource use parameters. Information about the identified studies is listed in Table 80 along with cross-references to tables reporting the identified values.

5.5.1.2 Details of included cost and resource use studies

Table 80: Details of cost and resource use studies

Study	Country	Date of study	Applicability to clinical practice in England	Costs for use in the CEM	Values from study
MTA ID 537 - Kobelt [116]	The UK	2002	High (utilised data from the ERAS study)	Indirect HAQ-related costs for inpatient care and joint replacement	See Table 85
MTA ID 537 - Wiles [117]	The UK	2005	Unclear	Direct HAQ-related costs for inpatient care and joint replacement	See Table 85
MTA ID 537 - Malottki [90]	The UK	2011	Unclear	Frequency of hospital outpatient attendance	10 outpatient visits during first 6 months of treatment, thereafter monthly.
Chakravarty [118]	The UK	2008	High	Frequency of monitoring (laboratory tests and other examinations)	See Table 84

5.5.1.3 Costing of resource use

The clinical management of RA requires frequent contact with health services including outpatient clinic visits, rheumatologist visits and regular laboratory testing. Patients may be occasionally hospitalised. Unit costs for these elements are available from the latest NHS Reference Costs and PSSRU Unit Costs of Health and Social Care. However, due to little evidence within the UK detailing frequency of resource use relating to RA, a range of sources and assumptions had to be used. Further details on these are provided in the following sections.

A clinical expert was interviewed to assess the costing of palliative care and assumptions for the administration of subcutaneous drugs, along with other elements not directly associated with costs and resource use. Table 81 provides further details related to the interview.

Table 81: Details of clinical expert recruitment and interview process

Item	Details
The criteria for selecting the experts	Rheumatologist with expert knowledge of treatment and palliative care of patients with RA in England
The number of experts approached	One
The number of experts who participated	One
Declaration of potential conflict(s) of interest from each expert whose opinion was sought	None recorded
The background information provided and its consistency with all the evidence provided in the submission	Questions were sent in advance, together with an explanation of the purpose of the interview
The method used to collect the opinions	Transcription of telephone interview
The medium used to collect opinions (for example, was information gathered by direct interview, telephone interview or self-administered questionnaire?)	Telephone interview
The questions asked	See Appendix 8.14.1.7
Whether iteration was used in the collation of opinions and if so, how it was used (for example, the Delphi technique)	Not relevant, only one expert interviewed

5.5.2 Intervention and comparators' costs and resource use

Most resource use and unit costs reflect latest costs in the British National Formulary and the latest unit costs of health and social care resources. Unit costs are reported in GBP (£) based on a 2015 cost year. Costs retrieved from earlier studies were inflated to 2015 using the health component of the UK consumer price index.[119]

5.5.2.1 Drug acquisition costs

The costs of drug acquisition were based on the recommended dosing schedules for treatment multiplied by the unit cost of treatment as reported in the British National Formulary 64.[22] The unit costs for each drug are displayed in Table 82.

Current recommendations on the use of CZP in England and Wales are subject to a PAS, in which treatment for patients is provided to the NHS free of charge for the first 12 weeks. The CZP PAS has been accounted for in the cost effectiveness analysis.

The PAS for GOL is a free stock arrangement, which provides the 100 mg dose at the same price as the 50 mg dose. The 100 mg dose is recommended for patients who weight more than 100 kg and fail to respond to three or four administrations of the 50 mg dose.

For all evaluations (basecase and sensitivity analyses), the costs of CZP and GOL are adjusted to take account of the PASs.

Table 82: Drug unit costs for bDMARDs

Treatment	Brand	Unit cost (2016 £)	Dose per unit (mg)	Dose description (SmPc)
CZP	Cimzia	£357.50	200.00	400 mg at week 0,2,4, and 200 mg Q2W thereafter
ABA (IV)	Orencia	£302.40	250.00	500-1000 mg (10 mg/kg) week 0,2,4 thereafter every 4 wks
ABA (SC)		£302.40	125.00	125 mg once per week
ADA	Humira	£352.14	40.00	40 mg every other week
ETA	Enbrel	£89.38	25.00	25 mg twice weekly
GOL	Simponi	£762.97	50.00	50 mg every 4 weeks
IFX	Remicade	£419.62	100.00	3 mg/kg week 0, 2 and 6 thereafter every 8 weeks
TOC (IV)	RoActemra	£256.00	200.00	8 mg/kg but no lower than 480 mg EO4W
TOC (SC)		£228.28	162.00	162 mg once per week
RTX	MabThera	£873.15	500.00	1000 mg wk 0 and 2, thereafter not more frequent then every 6 months
IFX	Inflectra or/ Remsima	£377.66	100.00	3 mg/kg week 0, 2 and 6 thereafter every 8 weeks

For IV drugs that are administered based on body weight (ABA, IFX, TOC, azathioprine and cyclosporine), the weight distribution of patients enrolled to the REALISTIC trial was applied to estimate the number of vials used.

For drugs that require loading doses or irregular administration, various assumptions were made to estimate the dose received by patients during the first and subsequent 6 months of treatment:

- For ABA (IV), it was assumed that during the first 6 months, treatment was administered at Weeks 0, 2, 4, 8, 12, 16, 20 and 24, equating to 8 administrations. During the subsequent 6 months, it was assumed that administrations occurred at a frequency of every 4 weeks, equating to 6.5 administrations over a 26-week cycle.
- For IFX (IV), similar assumptions were made when estimating dosing, where treatment was administered at Weeks 0, 2, 6, 14, and 22 during the first 6 months, and an average of 3.5 administrations during any subsequent 6-month period.
- For CZP, treatment was administered at Weeks 0, 2 and 4 during the first month of treatment, with further doses administered every two weeks on a continuous basis until cessation

The base case assume for drugs with IV administration, unopened vials are lost (ie. waste). Sensitivity analysis was performed assuming no waste (ie. unopened vials are consumed elsewhere in the system).

5.5.2.2 Drug administration and monitoring costs

The costs of administration and monitoring treatment programmes in patients with RA included costs associated with visits in outpatient settings for intravenous infusions, GP visits as well as certain monitoring tests and examinations. The calculated cost of monitoring treatment was estimated based on the unit costs shown in Table 83 and the monitoring schedules shown in Table 84. Frequency of laboratory tests and examinations was estimated from BSR Guidelines for DMARD therapy, published in 2008.[118] Since this report does not present data for all therapies in the current submission, the following assumptions had to be made. All biological therapies were assumed to have an identical monitoring schedule to MTX. The only exception was TOC, which required additional monitoring for neutrophils and platelets as well as lipid parameters, 4 to 8 weeks following initiation of therapy. The number of hospital outpatient attendance was based on values reported in a recent NICE appraisal (ID 537), which assumed 10 visits during the first 6 months and monthly visits thereafter, for all therapies.

The administration cost of intravenous infusions was retrieved from NICE TA247, estimated at £154 per infusion and then inflated to a present value of £174. This estimate accounts for 60 minutes infusion time which may be a favourable (ie underestimates) assumption for IFX, which has an infusion time of 2 hours, and unfavourable (ie overestimates) to ABA which takes 30 minutes to administer.

Table 83: Monitoring tests – unit costs

Item	Unit cost (2015 £)	Source
Rheumatologist visit	£137.00	NHS Reference Costs 14/15: WF01A
GP visit	£65.00	PSSRU 2015 (p. 177, 10.8b)
Nurse visit	£75.00	PSSRU 2015 (p. 172, 10.4)
Hospital day - Pal.	£371.00	PSSRU 2015 (p. 107, 7.1)
IV administration	£173.60	NICE TA247
Full blood count	£3.01	NHS Reference Costs 14/15: DAPS05
Urea and electrolytes	£1.19	NHS Reference Costs 14/15: DAPS04
Liver function test	£3.01	NHS Reference Costs 14/15: DAPS05
Creatinine	£3.01	NHS Reference Costs 14/15: DAPS05
Chest X-ray	£30.23	NHS Reference Costs 14/15: DAPF

Table 84: Monitoring assumptions

Treatment	Pre-treatment	First 6m on treatment	Subsequent 6m on treatment
MTX standard	FBC, U&E, LFT, CXR	11 x (FBC + U&E + LFT)	6.5 x (FBC + U&E + LFT)
Leflunomide	FBC, U&E, LFT, CRE	6.5 x (FBC + LFT)	3.25 x (FBC + LFT)
Ciclosporin	FBC, U&E, LFT, 2xCRE	6.5 x (FBC + LFT) 13 x (U&E + CRE)	2.16 x (FBC + LFT) 6.5 x (U&E + CRE)
Azathioprine	FBC, U&E, LFT	12 x (FBC + LFT), U&E + CRE	U&E, CRE
Sulfasalazine	FBC, U&E, LFT, CRE	4 x (FBC + LFT)	2.16 x (FBC + LFT)

FBC, full blood count; U&E, urea and electrolytes; LFT, liver function test; CXR, chest x-ray; CRE, creatinine. Hydroxychloroquine, MTX max dose, non-biologic and gold injections are assumed to have same monitoring schedule as MTX standard. CTZ, ABA, ADA, ETA, GOL, IFX, TOC and RTX also assumed to have same schedule as MTX standard.

5.5.3 Health-state unit costs and resource use

Total costs for each treatment are summarised in Appendix 8.14.2.8 broken down by direct costs which include drugs, administration and monitoring.

In addition to the direct medical costs associated with treatment, the cost effectiveness model also accounts for additional costs by HAQ-DI category, summarised in Table 85. This is considered a necessary part of the analysis since patients with more severe symptoms tend to have higher rates of hospitalisation and surgical procedures such as joint replacement. The costs are reported per HAQ band and estimated based on the Norfolk Arthritis Register (NOAR) database and NHS reference costs, which is consistent with the estimates used by the assessment group in a recent MTA (TA375). Indirect costs per HAQ band were retrieved from a study reporting costs from The Early RA Study (ERAS).[116] This study was selected in the absence of better evidence, despite less disease severity (baseline mean HAQ score of 1.11 and a mean symptom duration of 8.2 months) than is assumed in the base case economic analysis.

Table 85: Costs by HAQ-DI category

HAQ category	Direct costs (used in base case)		Total costs including indirect costs (used in sensitivity analyses)	
	Direct costs reported (2010 £)	Costs adjusted for currency and inflation (2015£)	Total costs reported (2001 US$)	Costs adjusted for currency and inflation (2016 £)
<0.6	£167.41	£188.72	$221	£189.62
0.6 - 1.1	£102.54	£115.59	$3767	£3,232.09
1.1 - 1.6	£364.68	£411.10	$5,185	£4,448.73
1.6 - 2.1	£523.68	£590.34	$7,910	£6,786.78
2.1 - 2.6	£1,246.26	£1,404.89	$12,045	£10,334.61
≥2.6	£2,687.97	£3,030.10	$12,548	£10,766.18

Exchange rate applied £1.00 = $1.58 (Q4 2015 average)

5.5.4 Adverse reaction unit costs and resource use

Since the safety profile of CZP is comparable to that of other bDMARDs, the costs of AEs and SAEs were excluded from the analysis.

5.5.5 Miscellaneous unit costs and resource use

The resource use for palliative care was estimated based on expert opinion following consultation with a rheumatologist. The expert rheumatologist estimated that patients receiving palliative care would require outpatient consultations at least every two months and treatment with intravenous prednisolone at least three times per year. It is expected that administration of intravenous prednisolone would require admission to a day-case setting. Given these assumptions, the total cost per 6 month cycle of palliative care is estimated at £978.

5.6 *Summary of base-case de novo analysis inputs and assumptions*

Values of model inputs and assumption made in the base-case de novo analysis are tabulated and discussed in the following sections.

5.6.1 Summary of base-case de novo analysis inputs

A list of all the variables included in the model is presented in Table 86.

Table 86: Summary of variables applied in the economic model

See next page

Variable	Value		Source/comment	
Population characteristics	**Mean**	**SE**	**Source/comment**	
Age	[redacted]	[redacted]	REALISTIC	
Gender (% female)	[redacted]	[redacted]	REALISTIC	
Baseline HAQ	[redacted]	[redacted]	REALISTIC	
Baseline EQ-5D	[redacted]	[redacted]	PREDICT	
Baseline pain	[redacted]	[redacted]	REALISTIC	
Weight distribution	**Weight (kg)**	**Distribution**	**Source/comment**	
	0 - 39.9 kg	0.00%	REALISTIC	
	40 - 44.9 kg	0.76%		
	45 - 49.9 kg	1.52%		
	50 - 54.9 kg	3.79%		
	55 - 59.9 kg	7.58%		
	60 - 64.9 kg	8.84%		
	65 - 69.9 kg	11.62%		
	70 - 74.9 kg	9.60%		
	75 - 79.9 kg	8.84%		
	80 - 84.9 kg	9.34%		
	85 - 89.9 kg	7.07%		
	90 - 94.9 kg	6.57%		
	95 - 99.9 kg	4.55%		
	100 - 104.9 kg	4.29%		
	105 - 109.9 kg	2.53%		
	110 - 114.9 kg	4.29%		
	115 - 119.9 kg	2.78%		
	120 - 200 kg	6.06%		
	Total	100.00%		
	Average weight	83.8kg		
Treatment efficacy of first therapy (response and subsequent HAQ change)				
Initial response (EULAR) to therapy	**No response**	**Moderate**	**Good**	**Source/comment**
CZP + MTX	[redacted]	[redacted]	[redacted]	
ABA + MTX	[redacted]	[redacted]	[redacted]	
ADA + MTX	[redacted]	[redacted]	[redacted]	
ETA + MTX	[redacted]	[redacted]	[redacted]	
GOL + MTX	[redacted]	[redacted]	[redacted]	
IFX + MTX	[redacted]	[redacted]	[redacted]	
TOC + MTX	[redacted]	[redacted]	[redacted]	
RTX + MTX	[redacted]	[redacted]	[redacted]	
Biosimiliar IFX + MTX	[redacted]	[redacted]	[redacted]	
CZP	[redacted]	[redacted]	[redacted]	
ADA	[redacted]	[redacted]	[redacted]	
ETA	[redacted]	[redacted]	[redacted]	
TOC	[redacted]	[redacted]	[redacted]	
Subsequent 6 month change in HAQ score	**Mean**	**SE**	**Source/comment**	
All treatments	[redacted]	[redacted]	REALISTIC trial - Observed data to Week 12,	

			and extrapolated (linear) from Week 12 to 24
Efficacy of second therapy (initial and subsequent 6 month HAQ change)			
First 6 month change in HAQ score	**Mean**	**SE**	**Source/comment**
Biologics	-0.39	0.04	Based on results of RADIATE study, suggesting -0.39 change for TOC + MTX. All other biologics assumed to be associated with same effect on HAQ (simplifying assumption)
cDMARDs	-0.05	0.01	RADIATE study reports change in HAQ for PBO + MTX of -0.05 (N=158)
Subsequent 6 month change in HAQ score	**Mean**	**SE**	**Source/comments**
Biologics	0.00	0.00	Assumption: no change in HAQ over time
Non-biologics	+0.02	0.00	Based on data from previous STAs, indicating 0.045 point increase in HAQ per year
Palliative care	+0.03	0.00	Based on data from previous STAs, indicating 0.06 point increase in HAQ per year
Duration of therapy			
First treatment	**Weibull scale (SE)**	**Weibull shape (SE)**	**Source/comments**
All treatment	0.4416 (0.0096)	0.7008 (0.0034)	NICE TA 195
Subsequent treatment – first 6 month probability of discontinuation	**Mean**	**SE**	**Source/comments**
CZP+ MTX	[redacted]	[redacted]	Based on treatment effect combined with RADIATE study baseline effects
ABA + MTX	[redacted]	[redacted]	
ADA + MTX	[redacted]	[redacted]	
ETA + MTX	[redacted]	[redacted]	
GOL + MTX	[redacted]	[redacted]	
IFX + MTX	[redacted]	[redacted]	
TOC + MTX	[redacted]	[redacted]	
RTX + MTX	[redacted]	[redacted]	
Biosimilar IFX+ MTX	[redacted]	[redacted]	
CZP	[redacted]	[redacted]	Based on CZP study baseline-trial effects combined with effect of treatment versus CZP
ADA	[redacted]	[redacted]	
ETA	[redacted]	[redacted]	
TOC	[redacted]	[redacted]	
cDMARDs	83.7%	2.9%	Based on RADIATE study baseline-trial effects where over 50% of population had received two or more TNFis - assumes equivalent efficacy between cDMARDs
Non-biologic	18.9%	-	Based on long-term retention data from Edwards et al. Assumption same risk of long-term discontinuation
Palliative care	0.0%	-	Assumption
Subsequent treatment – subsequent 6 month probability of discontinuation	**Mean**	**SE**	
Biologics	15.6%	3.1%	Based on Weibull model fitted to BSRBR data from TA195, assumption of no difference in long-

			term discontinuation rate by therapy
MTX standard, MTX max dose, M+H+S, Leflunomide	3.8%	0.8%	Based on long-term retention data from Edwards et al. Assumption same risk of long-term discontinuation
Gold injection	11.3%	2.3%	
Ciclosporine	7.2%	1.4%	
Azathioprine	6.0%	1.2%	
Non-biologic	3.8%	-	
Palliative care	0.0%	-	
Change in EQ-5D by response status			
EULAR response category	**Mean**	**SE**	**Source/comments**
No response	[redacted]	[redacted]	REALISTIC
Moderate response	[redacted]	[redacted]	REALISTIC
Good response	[redacted]	[redacted]	REALISTIC
% in utility during first 4 weeks of treatment	100%	0.10	Assumption
HAQ to EQ-5D (parameters for mapping algorithm from Hernandez et al.)			
Parameter	**Mean**	**SE**	**Source/comments**
Change in HAQ	-0.2102	0.0000	Hernandez Alava et al [115]
Change in pain	0.0000	0.0000	
HAQ and pain	[redacted]	[redacted]	
Mortality			
Rate ratio per unit HAQ	1.43	0.15	Norton et al [111]
Cost input data			
Drug acquisition	**Cost**	**mg per unit / units of drug**	**Comments** **All costs taken from the latest BNF**
CZP	£357.50	200.00 / 1	200mg vial - fixed dose
ABA	£302.40	250.00 / 1	Body weight dependent
ADA	£352.14	40.00 / 1	40mg vial - fixed dose
ETA	£89.38	25.00 / 1	25mg vial - fixed dose
GOL	£762.97	50.00 / 1	50mg vial - fixed dose
IFX (Remicade)	£419.62	100.00 / 1	Body weight dependent
TOC	£256.00	200.00 / 1	Body weight dependent
RTX	£873.15	500.00 / 1	500mg vials - fixed dose
Biosimilar IFX (Inflectra, Remsima)	£377.66	100.00 / 1	Body weight dependent
MTX standard	£2.40	2.50 / 24	2.5mg tablet
MTX max dose	£2.40	2.50 / 24	2.5mg tablet
Hydroxychloroquine	£5.31	200.00 / 60	200mg tablet
Sulfasalazine	£12.78	500.00 / 112	500mg tablet
Leflunomide	£10.18	10.00 / 30	10mg tablet
Gold injection	£4.56	10.00 / 1	10mg vial - fixed dose
Ciclosporin	£25.50	50.00 / 30	50mg tablet
Azathioprine	£3.24	25.00 / 28	25mg tablet
Prednisolone	£6.87	25.00 / 1	25mg vial - fixed dose
Other costs	**Value**	**Source/comments**	
Drug costing	Vial waste	(ie. Waste = unopened vials are lost, no waste = unopened vials are	

method		consumed elsewhere in the system)
Cost of administration	£173.60	MTA TA375 – inflated to 2015 £
Disease related direct cost (HAQ related)		
HAQ score	**Cost**	**Source/comments**
< 0.6	£188.72	Estimated based on data from the NOAR database,[117] and multiplied by NHS reference costs
0.6 < 1.1	£115.59	
1.1 < 1.6	£411.10	
1.6 < 2.1	£590.34	
2.1 < 2.6	£1,404.89	
≥ 2.6	£3,030.10	

5.6.2 Assumptions

The key assumptions of the analysis are as follows:

- The following assumptions were made during the analysis of initial response to therapy:
    - **All populations:** The comparative efficacy of CZP versus PBO was estimated by comparing the reported EULAR response probabilities for CZP at six months to response probabilities for PBO derived by mapping response at three months to response at six months
    - **Populations B/C:** Due to the lack of published trial data, ADA, ETA, and IFX were modelled as a treatment class assuming equivalent efficacy to GOL, when given in combination with MTX
    - **Populations B:** Biosimiliars to IFX were assumed to have equivalent efficacy to IFX
    - **Populations C:** The efficacy of TOC monotherapy was modelled assuming a consistent relative effect for TOC versus PBO and other therapies, when given as a combination or monotherapy regimen
    - **Populations C**: Due to the lack of published trial data, the efficacy of comparator TNFis was modelled as a treatment class. The efficacy of ADA and ETA monotherapy were modelled using the effect size estimates for GOL versus CZP, derived from the NMA of bDMARDs, given in combination with MTX
- EULAR moderate and good response at six months is prognostic of improvements in EQ-5D utility, and is routinely used as a stopping rule for biologic treatment in TNFi-IR.
- After the first six months of first treatment, the mean HAQ, pain and utility score of patients treated with bDMARDs is assumed to remain constant up to the point of treatment discontinuation

- Upon discontinuation of first treatment, patients are assigned a decrease in utility equal to that applied for the initial response to treatment and, immediately, an increase in utility resulting from response to subsequent treatment.
- Response to subsequent treatment is modelled based on the change in HAQ score, which is mapped to the change in EQ-5D utility using published mapping algorithms
- After the first six months of subsequent treatment, the mean HAQ of patients treated with bDMARD is assumed to remain constant up to treatment discontinuation. For patients treated with cDMARDs or palliative care, the mean HAQ score is assumed to increase at annual rates of 0.045 and 0.06 respectively. In the base case, the maximum mean HAQ score in the model is set to 2.76 (upper limit for HAQ is 3.0).
- In the base case, the costs of ABA and TOC were calculated based on the intravenous formulations of each respective drug, on the basis that the clinical efficacy of these therapies were derived from studies where therapy was given via infusion. The subcutaneous formulations of these therapies are considered in the model scenario analyses (see Table 101)
- In the base case, RTX re-treatment occurs at a frequency of once every six months
- Based on clinical expert opinion it was assumed that patients receiving palliative care require routine assessments with a rheumatologist at a frequency of once every two months, and require ongoing treatment with intravenous prednisolone in a day case setting at a frequency of once every four months
- Treatment monitoring schedule for hydroxychloroquine, MTX max dose, non-biologic and gold injections were assumed equal to MTX standard therapy. Similarly, the monitoring frequency for all biologics was assumed to be the same as for MTX standard therapy.

5.7 *Base-case results*

The base-case results are summarised in the following sections, and include overall results and a full breakdown of costs, QALYs and life-years. Full probabilistic and deterministic sensitivity analyses are provided in addition to a number of scenario analyses.

5.7.1 Base-case incremental cost effectiveness analysis results

The results of the deterministic analysis show that

- **Population A**: Treatment with CZP followed by RTX and standard therapies is more expensive (+£9,938 per patient) and more effective (+0.288 QALYs

per patient) than a sequence of RTX followed by standard therapies, with a deterministic ICER of £34,516 per QALY gained (Table 87).

- **Population B**: Treatment sequences starting with either GOL or ADA are the least costly and least effective sequences in the analysis, and are considered to be an optimal treatment strategies for thresholds of up to £3,641 per QALY gained. At thresholds of between £3,641 and £129,319, CZP is the optimal treatment strategy (Table 88). At thresholds of £129,319 or greater, the TOC treatment sequence is the optimal strategy. The ETA, IFX and ABA treatment sequences were not considered in the fully incremental analysis as these strategies were dominated by GOL and ADA (ETA and IFX), and dominated by CZP (ABA).
- **Population C**: The ADA treatment sequence was the least-costly and less effective sequence in the analysis, and is considered the optimal treatment strategy for thresholds of up to £4,985 per QALY gained. At thresholds of between £4,985 and £123,915, CZP is the optimal treatment strategy (Table 89). At thresholds of £123,915 or greater, TOC is the optimal strategy. The ETA sequence was dominated (more costly and of equal efficacy) by ADA and not considered in the full incremental analysis.

Table 87: Base-case cost effectiveness results in population A (deterministic)

Technologies (branded biosimilar)	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER (£/QALY) versus baseline	ICER (£/QALY) incremental
RTX + MTX (MabThera)	£139,524	16.132	6.975					Cost effective at WTP < £34,516
CZP+ MTX (Cimzia)	£149,462	16.237	7.263	£9,938	0.105	0.288	£34,516	**Cost effective at WTP > £34,516**
ICER, incremental cost effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life years								

Table 88: Base-case cost effectiveness results in population B (deterministic, ordered in terms of least to most expensive)

Technologies	Total costs (£)	Total LYG	Total QALYs	Incremental costs versus baseline (£)	Incremental LYG versus baseline	Incremental QALYs versus baseline	ICER (£/QALY) versus baseline	ICER (£/QALY) incremental
GOL + MTX (Simponi)	£97,593	15.881	6.016					**Cost effective at WTP < £3,641**
ADA + MTX (Humira)	£97,593	15.881	6.016	£0	0.000	0.000	-	
ETA + MTX (Enbrel)	£98,017	15.881	6.016	£423	0.000	0.000	Dominated by ADA/GOL	Dominated by ADA/GOL
CZP + MTX (Cimzia)	£98,575	15.922	6.286	£981	0.041	0.270	£3,641	**Cost effective at WTP between £3,641 and £129,319**
IFX + MTX (Remicade)	£101,894	15.881	6.016	£4,300	0.000	0.000	Dominated by ADA/GOL	Dominated by ADA/GOL
ABA + MTX (IV - no PAS) (Orencia)	£115,609	15.889	6.065	£18,016	0.007	0.049	£370,920	Dominated by CZP
TOC + MTX (IV - no PAS) (RoActemra)	£125,096	15.953	6.491	£27,503	0.072	0.475	£57,946	**Cost effective at WTP > £129,319**
ICER, incremental cost effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life years								

Table 89: Base-case cost effectiveness results in population C (deterministic)

Technologies	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER (£/QALY) versus baseline	ICER (£/QALY) incremental
ADA (Humira)	£95,943	15.844	5.845					**Cost effective at WTP < £4,985**
ETA (Enbrel)	£96,347	15.844	5.845	£404	0.000	0.000	Dominated by ADA	Dominated by ADA
CZP (Cimzia)	£97,292	15.887	6.115	£1,349	0.043	0.271	£4,985	**Cost effective at WTP between £4,985 and £123,915**
TOC (IV – no PAS) (RoActemra)	£123,695	15.920	6.328	£27,752	0.076	0.484	£57,375	**Cost effective at WTP > £123,915**
ICER, incremental cost effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life years								

5.7.2 Clinical outcomes from the model

The clinical outcomes from the model are presented in terms of the time spent (undiscounted) in each state in the model.

A summary of the time spent in each state for the CZP and RTX sequences of population A is presented in Appendix 8.14.4.1.

In population A, it is estimated that patients who receive RTX will spend an additional 0.059 years on their first therapy, when compared to patients who receive CZP. Due to the additional lines of therapy in the CZP sequence (CZP followed by RTX and standard therapies versus RTX followed by standard therapies), patients in the CZP sequence spend an additional 0.304 years on subsequent therapy, compared to patients who receive the RTX sequence. As such, treatment with CZP is associated with a net gain in life expectancy of 0.245 years (or 0.105 when discounted at 3.5% per annum), versus the RTX sequence. The modelled gain in life expectancy for CZP is driven by the prolonged use of biologic therapies (four lines of biologic therapy with CZP versus three lines of biologic for RTX), which extends the time spent at lower HAQ scores and reduces the risks of excess mortality from RA.

Appendix
8.14.4.2 provides a summary of the mean time spent in each health state of the model, for cohorts allocated to CZP, ABA, comparator TNFi (ADA, GOL, ETA and IFX) and TOC, in population B of the appraisal.

The clinical outcomes for all comparator TNFis are presented as a single group as these therapies are assumed to have the same treatment effect in the model, and therefore generate identical results for the time spent in each model state.

In population B, the mean time spent on first therapy varied from approximately 3.5 years for comparator TNFi to 4.5 years for TOC (mean time spent on CZP was 4.0 years). After first therapy, the mean time on subsequent therapy ranged from 19.467 (TOC) to 20.266 years (comparator TNFi). The model predicts that RA patients would spend the majority of this period in the palliative care state (12.242 to 12.916 years), having exhausted all available treatment options (Appendix 8.14.4.2).

Appendix 8.14.4.3 provides a summary of the mean time spent in each health state of the model, for cohorts allocated to CZP, comparator TNFi (ADA, and ETA) and TOC monotherapy sequences in population C

The mean time spent in each state in the model in population C is broadly consistent with the clinical outcomes presented for population B. However, due to the limited number of treatment options available in patients unable to tolerate MTX, it is estimate that, on average, 14 of the 20 years spent on subsequent therapy will be spent receiving palliative care.

5.7.3 Disaggregated results of the base case incremental cost effectiveness analysis

A summary of disaggregated QALYs by health state for the CZP and RTX sequences of population A, is presented in Table 90.

In population A, the probability of EULAR response for RTX is numerically greater than the corresponding response probability for CZP (see section 5.3.1), which leads to a high number of moderate and good responders in the RTX arm of the model. As such, treatment with RTX is associated with a gain in QALYs versus CZP, when directly comparing the outcomes of initial therapy (incremental QALY of -0.042 for CZP versus RTX at first therapy). After failure on initial therapy, there is however differences in the sequence of treatments available to each cohort in the model, with a greater number of effective treatment options being available in the CZP sequence. In this scenario, the model predicts that follow on therapies to CZP will provide a net QALY gain when compared to the subsequent therapies given after RTX. The net gain in QALYs associated with subsequent therapies to CZP is +0.330 per patient-lifetime. Overall, the net QALY gains from using CZP prior to a standard RTX sequence is estimated at +0.288. In terms of percentage absolute increment, it is estimated that 97% of the difference in QALYs between groups is related to subsequent therapy use.

Table 90: Summary of discounted QALYs by health state – population A – per patient

Health state	CZP + MTX	RTX + MTX	Increment	Absolute increment	% absolute increment
First therapy					
No response	0.049	0.046	0.003	0.003	0.12%
Moderate response	1.274	1.260	0.014	0.014	0.51%
Good response	1.126	1.185	-0.059	0.059	2.18%
Subtotal	2.450	2.492	-0.042	0.077	2.82%
Subsequent therapy					
2nd therapy	0.801	0.939	-0.138	0.138	5.09%
3rd therapy	0.879	0.585	0.294	0.294	10.83%
4th therapy	0.537	0.441	0.095	0.095	3.52%
5th therapy	0.395	1.411	-1.016	1.016	37.42%
6th therapy	1.251	1.107	0.144	0.144	5.32%
7th therapy	0.950	0.000	0.950	0.950	35.00%
Subtotal for FU therapies	4.813	4.483	0.330	2.638	97.18%
All therapies					
Total (all therapies)	7.263	6.975	0.288	2.715	100.00%

FU, follow-up
Adapted from Pharmaceutical Benefits Advisory Committee (2008) Guidelines for preparing submissions to the Pharmaceutical Benefits Advisory Committee (Version 4.3). Canberra: Pharmaceutical Benefits Advisory Committee

A summary of disaggregated QALYs for the CZP, ABA, comparator TNFi (ADA, GOL, ETA and IFX) and TOC sequences for population B and population C is presented in Appendix 8.14.4.2 and 8.14.4.3.

In population B, treatment with CZP is associated with an incremental QALY gain of 0.221 versus ABA, of 0.270 versus comparator TNFi, and a net QALY loss of 0.205 versus TOC. Across all comparisons in population B, it is estimated that over 80% of the percentage absolute increment in QALYs is associated with response to first therapy, with fewer than 20% of QALY increments coming from subsequent therapy use.

In population C, treatment with CZP is associated with net QALY gains of 0.271 versus the comparator TNFi, and a net QALY loss of 0.213 versus TOC. As with population B, over 80% of the percentage absolute increment in QALYs in population C is associated with response to first therapy.

Table 91: Summary of discounted QALYs by health state – population B – per patient

Health state	Absolute QALY				Increment QALY (CZP vs.)			% absolute increment QALY (CZP vs.)		
	CZP + MTX	ABA + MTX	Comparator TNF*	TOC + MTX	ABA + MTX	Comparator TNF*	TOC + MTX	ABA + MTX	comparator TNF*	TOC + MTX
First therapy										
No response	0.049	0.075	0.081	0.028	-0.026	-0.032	0.021	5.13%	5.43%	2.99%
Moderate response	1.274	1.299	1.289	1.116	-0.025	-0.015	0.159	4.87%	2.46%	22.14%
Good response	1.126	0.762	0.696	1.587	0.364	0.430	-0.461	71.78%	72.83%	64.31%
Subtotal	2.450	2.136	2.066	2.730	0.313	0.383	-0.281	81.78%	80.71%	89.44%
2nd therapy	0.522	0.529	0.531	0.516	-0.007	-0.009	0.006	1.45%	1.54%	0.83%
3rd therapy	0.470	0.477	0.479	0.465	-0.007	-0.009	0.006	1.37%	1.45%	0.79%
4th therapy	0.294	0.300	0.302	0.289	-0.006	-0.008	0.005	1.24%	1.31%	0.72%
5th therapy	0.323	0.330	0.332	0.317	-0.007	-0.009	0.006	1.44%	1.52%	0.84%
6th therapy	0.313	0.320	0.322	0.307	-0.007	-0.009	0.006	1.46%	1.54%	0.85%
7th therapy	1.914	1.971	1.985	1.867	-0.057	-0.070	0.047	11.26%	11.93%	6.54%
Subtotal for FU therapies	3.836	3.929	3.950	3.760	-0.092	-0.114	0.076	18.22%	19.29%	10.56%
All therapies										
Total (all therapies)	6.286	6.065	6.016	6.491	0.221	0.270	-0.205	100.00%	100.00%	100.00%

FU, follow-up

Adapted from Pharmaceutical Benefits Advisory Committee (2008) Guidelines for preparing submissions to the Pharmaceutical Benefits Advisory Committee (Version 4.3). Canberra: Pharmaceutical Benefits Advisory Committee

Note: * results grouped for ADA, IFX and ETA, as they are assumed to have the same effect in the model

Table 92: Summary of discounted QALYs by health state – population C – per patient

Health state	Absolute QALY			Increment QALY (CZP vs.)		% absolute increment QALY (CZP vs.)	
	CZP	Comparator TNF*	TOC	comparatorTNF*	TOC	Comparator TNF*	TOC
First therapy							
No response	0.057	0.091	0.033	-0.034	0.024	5.90%	4.01%
Moderate response	1.533	1.436	1.450	0.096	0.083	16.69%	14.03%
Good response	0.734	0.406	1.135	0.327	-0.402	56.79%	68.03%
Subtotal	2.323	1.933	2.618	0.389	-0.295	79.39%	86.07%
Subsequent therapy							
2nd therapy	0.524	0.534	0.517	-0.010	0.007	1.67%	1.12%
3rd therapy	0.320	0.323	0.317	-0.004	0.003	0.64%	0.43%
4th therapy	0.367	0.377	0.360	-0.010	0.007	1.82%	1.23%
5th therapy	0.357	0.367	0.349	-0.011	0.007	1.84%	1.25%
6th therapy	2.225	2.309	2.166	-0.084	0.059	14.63%	9.91%
Subtotal for FU therapies	3.792	3.911	3.710	-0.119	0.082	20.61%	13.93%
All therapies							
Total (all therapies)	6.115	5.845	6.328	0.271	-0.213	100.0%	100.0%

Note: * results grouped for ADA and ETA, as they are assumed to have the same effect in the model

Table 93 (population A), Table 94 and Table 95 (population B) and Table 96 (population C) provide summaries of the cost by health state in the model.

In population A, the CZP sequence was more expensive than the RTX sequence, with an incremental total cost of £9,938 per patient. Approximately 60% of the incremental costs for CZP were the result of subsequent therapy use, and in particular, the acquisition costs of subsequent drug therapies (40% of absolute drug costs). The next largest contributor to absolute incremental cost was the cost of drug acquisition for CZP, which was associated with a net cost of £5,250, when compared directly to the cost of RTX. These additional costs were however, partially offset by a reduction in the cost of intravenous administrations in the CZP sequence. The costs for drug administration, drug monitoring and hospitalisation accounted for between 10 and 15% of the absolute incremental costs shown in Table 93.

Table 93: Summary of costs by health state – population A – per patient

Health state	CZP + MTX	RTX + MTX	Increment	Absolute increment	% absolute increment
First treatment					
Drug	£30,193	£24,943	£5,250	£5,250	26.82%
Administration	£0	£2,469	-£2,469	£2,469	12.61%
Monitoring	£7,072	£7,165	-£93	£93	0.47%
Hospital costs (HAQ)	£2,119	£2,114	£5	£5	0.02%
Indirect costs (HAQ)	£0	£0	£0	£0	0.00%
Subtotal	£39,384	£36,690	£2,694	£7,816	39.93%
Subsequent therapy					
Drug	£53,867	£46,158	£7,708	£7,708	39.37%
Administration	£18,191	£18,584	-£392	£392	2.00%
Monitoring	£20,208	£18,414	£1,794	£1,794	9.16%
Hospital costs (HAQ)	£17,812	£19,678	-£1,866	£1,866	9.53%
Indirect costs (HAQ)	£0	£0	£0	£0	0.00%
Subtotal	£110,078	£102,834	£7,244	£11,760	60.07%
All treatments					
Total	£149,462	£139,524	£9,938	£19,577	100.00%

FU, follow-up

In population B, the CZP sequence was more costly than the other comparator TNFi sequences, with a total net cost of £981 per patient-lifetime. The total additional costs were largely driven by an increase in the cost of first therapy use in the CZP sequence (41.85% of absolute increment, net cost of £2,382), which is the result of patients spending longer on CZP than other TNFi therapies (see Appendix 8.14.4.2). Further, by delaying the switch to subsequent therapies, treatment with CZP was also associated with savings in terms of the costs of subsequent therapy. This included net cost-savings of £1,283 for hospitalisation costs, £585 for administration

costs and £371 for monitoring costs. In comparison to TOC (population B), the CZP sequence was cost-saving with an incremental total cost of -£26,521 (CZP versus TOC). The incremental costs of TOC were largely driven by the additional cost for the acquisition and administration of TOC (acquisition = -£19,086, 63.38% of absolute incremental, administration = £8,634, 28.67%).

Similar outcomes were observed in population C where the CZP sequence was found to be more costly than ADA and ETA (net total costs of £1,349 and £945 respectively), and cost-saving when compared to TOC (net cost-saving of £26,403). As reported in population B, the key cost driver in the analysis was drug acquisition for the first therapy in the sequence, which accounted for between 43% and 64% of the absolute incremental costs between sequences.

Table 94: Summary of costs by health state – population B - per patient

Health state	CZP + MTX	ABA + MTX	ADA + MTX	ETA + MTX	GOL + MTX	IFX + MTX	TOC + MTX
First treatment							
Drug	£30,193	£39,017	£27,811	£28,234	£27,811	£28,173	£49,279
Administration	£0	£7,294	£0	£0	£0	£3,938	£8,634
Monitoring	£7,072	£6,351	£6,184	£6,184	£6,184	£6,184	£7,669
Hospital costs (HAQ)	£2,119	£2,078	£2,053	£2,053	£2,053	£2,053	£2,023
Indirect costs (HAQ)	£0	£0	£0	£0	£0	£0	£0
Subtotal	£39,384	£54,740	£36,048	£36,471	£36,048	£40,348	£67,605
Subsequent therapy							
Drug	£3,714	£3,809	£3,831	£3,831	£3,831	£3,831	£3,637
Administration	£18,118	£18,592	£18,703	£18,703	£18,703	£18,703	£17,730
Monitoring	£13,084	£13,385	£13,455	£13,455	£13,455	£13,455	£12,838
Hospital costs (HAQ)	£24,274	£25,083	£25,557	£25,557	£25,557	£25,557	£23,286
Indirect costs (HAQ)	£0	£0	£0	£0	£0	£0	£0
Subtotal	£59,190	£60,869	£61,546	£61,546	£61,546	£61,546	£57,491
All treatments							
Total	£98,575	£115,609	£97,593	£98,017	£97,593	£101,894	£125,096

FU, follow-up

Table 95: Summary of costs by health state – population B (after excluding dominated and extendedly dominated therapies) - per patient

	Increment (CZP. Vs)		Absolute increment (CZP vs.)		% absolute increment (CZP vs.)	
Health state	**GOL + MTX / ADA + MTX**	**TOC + MTX**	**GOL + MTX / ADA + MTX**	**TOC + MTX**	**GOL + MTX / ADA + MTX**	**TOC + MTX**
First treatment						
Drug	£2,382	-£19,086	£2,382	£19,086	41.85%	63.38%
Administration	£0	-£8,634	£0	£8,634	0.00%	28.67%
Monitoring	£888	-£596	£888	£596	15.61%	1.98%
Hospital costs (HAQ)	£66	£96	£66	£96	1.16%	0.32%
Indirect costs (HAQ)	£0	£0	£0	£0	0.00%	0.00%
Subtotal	£3,337	-£28,220	£3,337	£28,413	58.62%	94.36%
Subsequent therapy						
Drug	-£116	£77	£116	£77	2.05%	0.26%
Administration	-£585	£388	£585	£388	10.28%	1.29%
Monitoring	-£371	£246	£371	£246	6.52%	0.82%
Hospital costs (HAQ)	-£1,283	£988	£1,283	£988	22.54%	3.28%
Indirect costs (HAQ)	£0	£0	£0	£0	0.00%	0.00%
Subtotal	-£2,355	£1,699	£2,355	£1,699	41.38%	5.64%
All treatments						
Total	£981	-£26,521	£5,692	£30,112	100.00%	100.00%

Table 96: Summary of costs by health state – population C (increments excluding ETA as extendedly dominated) - per patient

Health state	Absolute				Increment (CZP. Vs)		Absolute increment (CZP vs.)		% absolute increment (CZP vs.)	
	CZP	ADA	ETA	TOC	ADA	TOC	ADA	TOC	ADA	TOC
First treatment										
Drug	£29,080	£26,458	£26,863	£48,199	£2,621	-£19,120	£2,621	£19,120	42.75%	63.56%
Administration	£0	£0	£0	£8,465	£0	-£8,465	£0	£8,465	0.00%	28.14%
Monitoring	£6,871	£5,929	£5,929	£7,530	£943	-£658	£943	£658	15.38%	2.19%
Hospital costs (HAQ)	£2,346	£2,170	£2,170	£2,327	£176	£19	£176	£19	2.87%	0.06%
Indirect costs (HAQ)	£0	£0	£0	£0	£0	£0	£0	£0	0.00%	0.00%
Subtotal	£38,297	£34,557	£34,961	£66,521	£3,740	-£28,224	£3,740	£28,262	61.00%	93.95%
Subsequent therapy										
Drug	£3,646	£3,761	£3,761	£3,567	-£114	£79	£114	£79	1.87%	0.26%
Administration	£20,315	£20,978	£20,978	£19,857	-£663	£457	£663	£457	10.81%	1.52%
Monitoring	£9,578	£9,861	£9,861	£9,383	-£283	£195	£283	£195	4.61%	0.65%
Hospital costs (HAQ)	£25,456	£26,787	£26,787	£24,366	-£1,331	£1,090	£1,331	£1,090	21.71%	3.62%
Indirect costs (HAQ)	£0	£0	£0	£0	£0	£0	£0	£0	0.00%	0.00%
Subtotal	£58,995	£61,386	£61,386	£57,174	-£2,391	£1,821	£2,391	£1,821	39.00%	6.05%
All treatments										
Total	£97,292	£95,943	£96,347	£123,695	£1,349	-£26,403	£6,131	£30,083	100.00%	100.00%

FU, follow-up

5.8 *Sensitivity analyses*

5.8.1 Probabilistic sensitivity analysis

Table 97 provides a summary of the model parameters that were sampled in the probabilistic sensitivity analysis. Cost parameters including drug acquisition, administration, and monitoring were excluded from the probabilistic analysis.

Table 97: Parameters varied through probabilistic sensitivity analyses

Parameter	Simulation	Source
Weight	Sampled using Dirichlet distribution. The sample size for the distribution was based on data from the REALISTIC trial	REALISTIC trial
Age	Normal distribution defined by the mean (54.27) and its SE (0.601)	REALISTIC trial
Gender	Beta distribution defined by N (396) and n (0.78*396)	REALISTIC trial
Baseline HAQ	Normal distribution defined by the mean (1.55) and its SE (0.032)	REALISTIC trial
Baseline EQ-5D	[redacted]	PREDICT trial
Baseline pain	Beta distribution defined by the mean (0.63) and its SE (0.012)	REALISTIC trial
Short-term response to treatment	Normal distributions defined by the values associated with treatment effect, response on reference as well as change in HAQ from the first 6 months on therapy	NMA
Long-term response to treatment	Normal distribution defined by the values associated with long-term response to treatment	REALISTIC trial
Long-term HAQ	Normal distribution defined by values associated with long-term HAQ score for each treatment	Assumption: no change in HAQ over time
HAQ to EQ-5D	Normal distributions defined by values for change in HAQ and pain, retrieved from the first class model published by Hernandez et al	Hernandez Alava et al [115]
Duration of therapy	Scale and shape parameters for the Weibull distribution varied using a normal distribution defined by the mean for shape and scale respectively (0.4416 and 0.7008) and their SEs (0.0096 and 0.0034) Normal distribution was used for treatment duration for subsequent treatments, defined by parameter estimates from a range of sources	NICE TA195 RADIATE study, Edwards et al [107]
HAQ mortality multiplier	Lognormal distribution defined by the mean (1.430) and its SE (1.43)	Norton et al [111]

The results of the probabilistic sensitivity analysis are based on a Monte Carlo simulation of 5,000 iterations. The total costs and total effectiveness of treatment were recorded at each simulation of the analysis, and saved to the models PSA sheet. Once the PSA is complete, estimates of the mean costs and mean QALYs for each treatment option in the model is generated, and used to calculate the probabilistic ICERs (difference in mean cost divided by difference in mean QALY).

The output of the analysis includes the probabilistic ICER, an assessment of the probability of cost effectiveness at thresholds of £20,000 and £30,000 per QALY

gained, and the generation of a series of multi-way/one-way cost effectiveness acceptability curves.

The results of the probabilistic sensitivity analysis are provided in Table 98, Table 99, and Table 100. Cost effectiveness acceptability curves are shown in Figure 70, Figure 71, and Figure 72.

When compared to the deterministic results, there were only minor differences in the mean costs and effectiveness generated in the probabilistic analysis. The percentage difference in mean results (probabilistic mean versus deterministic mean) ranged from 0.1 to 1.0% across all populations.

In population A, there was a 37.40% probability that CZP was cost effective versus RTX at a threshold of £30,000 per QALY. The mean ICER for this comparison was £33,665 per QALY gained. At a threshold of £20,000 per QALY gained, the probability of cost effectiveness reduced from 37.40% to 2.98%.

In populations B and C, there was a 95.9% (B) and 97.9% (C) probability that CZP is the optimal treatment strategy at a threshold of £30,000. At a lower threshold of £20,000, the probability that CZP is the optimal treatment strategy was in excess of 99% (99.50% - population B and 99.64% - population C, respectively).

Table 98. Probabilistic sensitivity analysis results – population A

Therapy	Mean costs	Difference in costs (CZP vs. treatment)	Mean QALYs	Difference in QALYs (CZP vs. treatment)	ICER (CZP vs. treatment)	Probability of cost effectiveness at WTP threshold of £20,000/QALY (%)	Probability of cost effectiveness at WTP threshold of £30,000/QALY (%)
CZP + MTX	£150,413	£9,782	7.295	0.291	£33,665	2.98%	37.40%
RTX + MTX	£140,631		7.005			97.02%	62.60%

Results represent mean estimates per patient

Table 99. Probabilistic sensitivity analysis results – population B

Therapy	Mean costs	Difference in costs (CZP vs. treatment)	Mean QALYs	Difference in QALYs (CZP vs. treatment)	ICER (CZP vs. treatment)	Probability of cost effectiveness at WTP threshold of £20,000/QALY (%)	Probability of cost effectiveness at WTP threshold of £30,000/QALY (%)
CZP + MTX	£98,916		6.302			99.5%	95.9%
ABA + MTX	£116,217	-£17,301	6.087	0.214	Cimzia is dominant	0.0%	0.0%
ADA + MTX	£97,944	£972	6.034	0.267	£3,635	0.2%	1.2%
ETA + MTX	£98,402	£513	6.037	0.265	£1,938	0.1%	1.1%
GOL + MTX	£97,984	£931	6.038	0.264	£3,527	0.2%	1.7%
IFX + MTX	£102,272	-£3,356	6.038	0.263	Cimzia is dominant	0.0%	0.0%
TOC + MTX	£125,518	-£26,603	6.507	-0.206	£129,321	0.0%	0.0%

**Original brands only (ie Remicade (IFX) and Enbrel (ETA))
CZP dominates comparator (ie. CZP is more effective and less costly)
Results represent mean estimates per patient

Table 100. Probabilistic sensitivity analysis results – population C

Therapy	Mean costs	Difference in costs (CZP vs. treatment)	Mean QALYs	Difference in QALYs (CZP vs. treatment)	ICER (CZP vs. treatment)	Probability of cost effectiveness at WTP threshold of £20,000/QALY	Probability of cost effectiveness at WTP threshold of £30,000/QALY (%)
CZP	£97,550		6.141			99.64%	97.20%
ADA	£96,198	£1,352	5.868	0.274	£4,943	0.240%	1.700%
ETA	£96,587	£963	5.867	0.274	£3,514	0.120%	1.100%
TOC	£123,749	-£26,199	6.344	-0.203	£129,177	0.000%	0.000%

* No patient access scheme was taken into account for TOC
**Original brands only (ie Enbrel (ETA))
Results represent mean estimates per patient

Figure 70. Cost effectiveness acceptability curves – population A

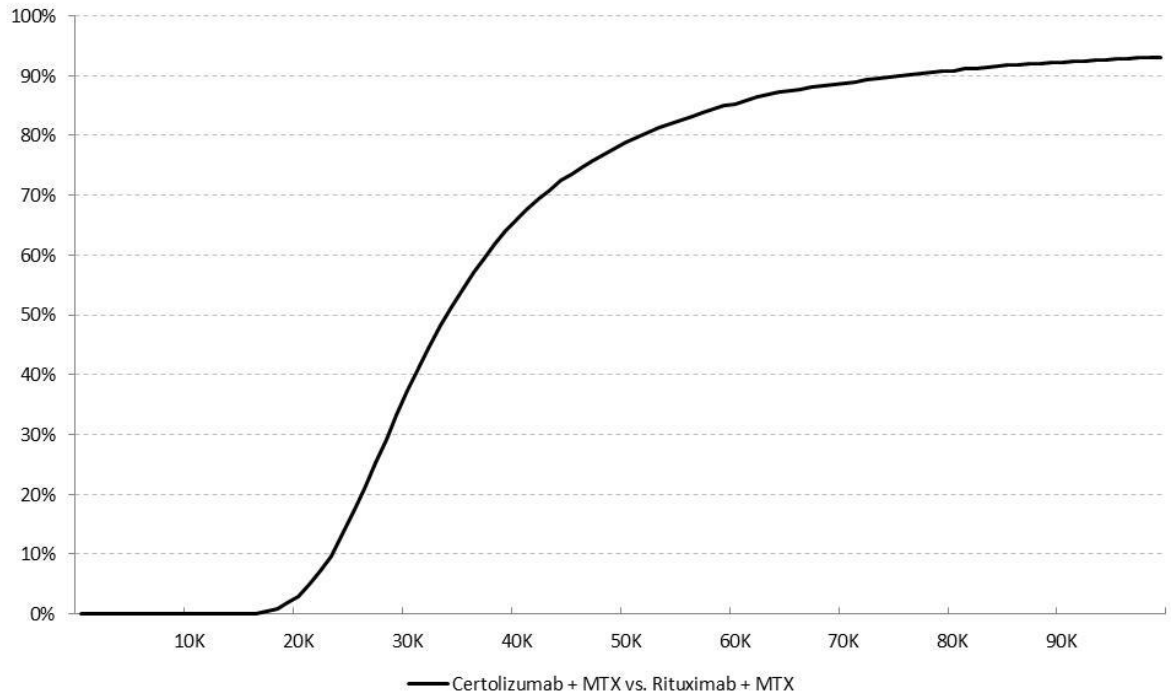


Figure 71. Cost effectiveness acceptability curves – population B

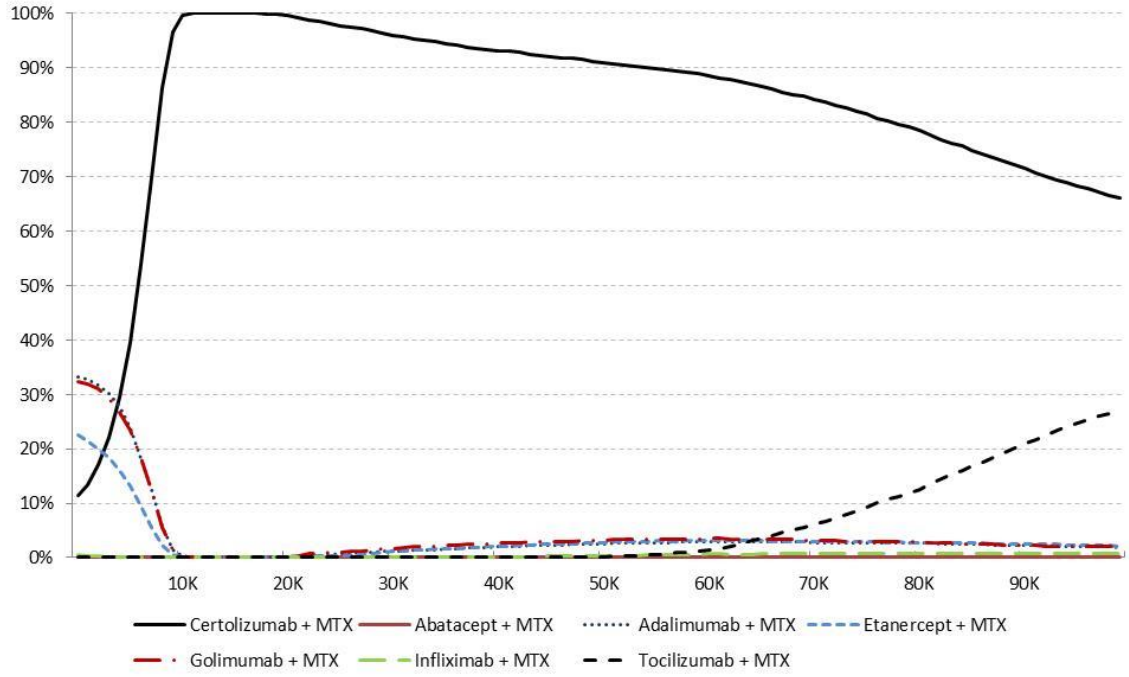
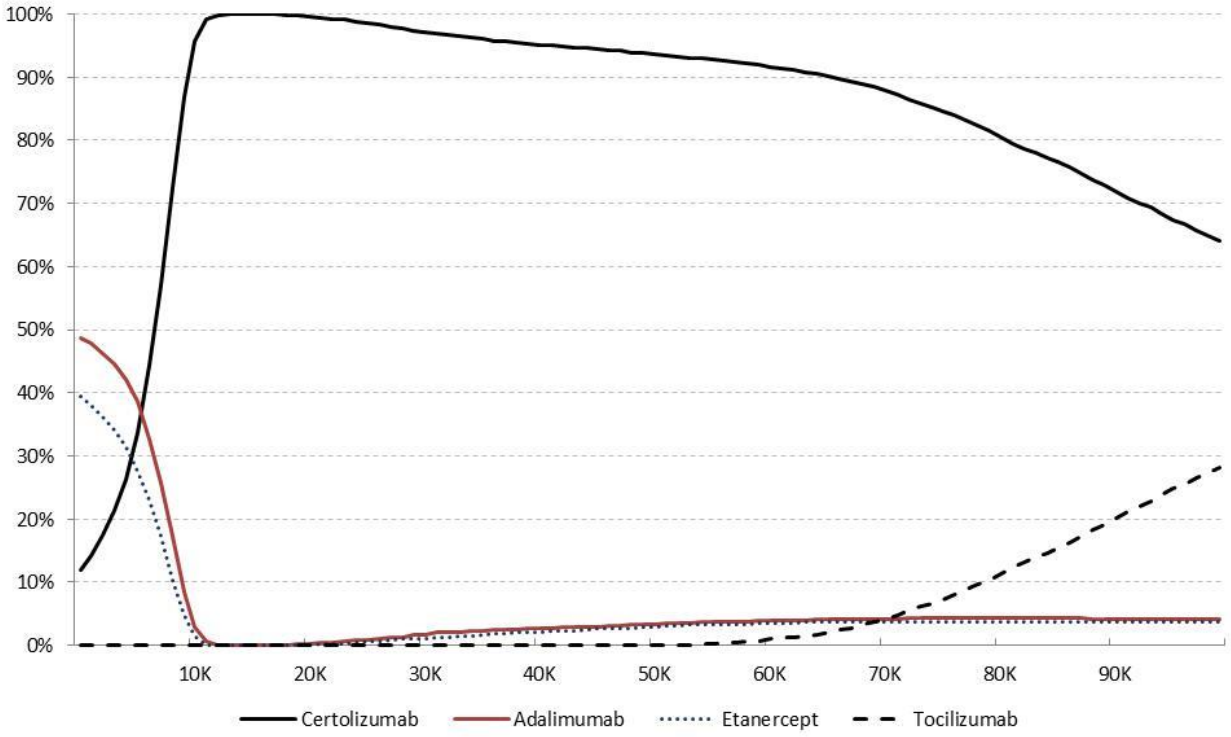


Figure 72. Cost effectiveness acceptability curves – population C



5.8.2 Deterministic sensitivity analysis

Deterministic sensitivity analyses were performed on comparators that were neither dominated nor extendedly dominated in the base case analysis. This includes:

- **Population A:**
  - CZP + MTX versus RTX + MTX
- **Population B:**
  - CZP + MTX versus ADA/GOL/TOC + MTX
- **Population C:**
  - CZP versus ADA/TOC

The parameters that were considered in the one-way deterministic sensitivity analysis comprise discount rates for costs and effects (0 to 6%), baseline HAQ, pain, and EQ-5D scores (fixed 30% variation), trial-specific baseline effects and cut-off statistics for the NMA model (based on 95% confidence interval), HAQ mortality multiplier (based on 95% confidence interval), relationship between change in HAQ, change in pain and EQ-5D (fixed 30% variation), and the effect of treatment (CZP or comparator) on EULAR response probability (based on 95% confidence interval).

The results of the sensitivity analysis are presented as tornado diagrams in Figure 73, Figure 74, and Figure 75. The x-axis on the graphs shows the percentage change in net monetary benefit at a willingness to pay threshold of £30,000.

Across all three populations (A, B and C), the parameters that exhibited the greatest influence on results were the comparative efficacy of CZP and comparator on EULAR response at six months, the discounting rates for costs and effects, and the change in EQ-5D per unit change in HAQ. Variation in mean baseline HAQ, pain and EQ-5D scores, the HAQ mortality multiplier, trial-specific baseline effects and the cut-off statistics for the NMA model were found to have a small impact on results.

In a number of scenarios, a 30% variation in mean baseline HAQ resulted in positive changes to net monetary benefit (ie. a higher ICER compared to the base case). This result was driven by non-linearities in the model introduced by the restriction of HAQ scores to between 0 and 3. The magnitude of change in net monetary benefit associated with mean baseline HAQ was modest, and therefore this parameter is not considered a key driver of the results of the cost effectiveness analysis.

Figure 73: Tornado diagram showing the percentage change in net monetary benefit at a willingness to pay threshold of £30,000 per QALY gained, based on variation in individual model parameters – population A (CZP + MTX versus RTX + MTX)

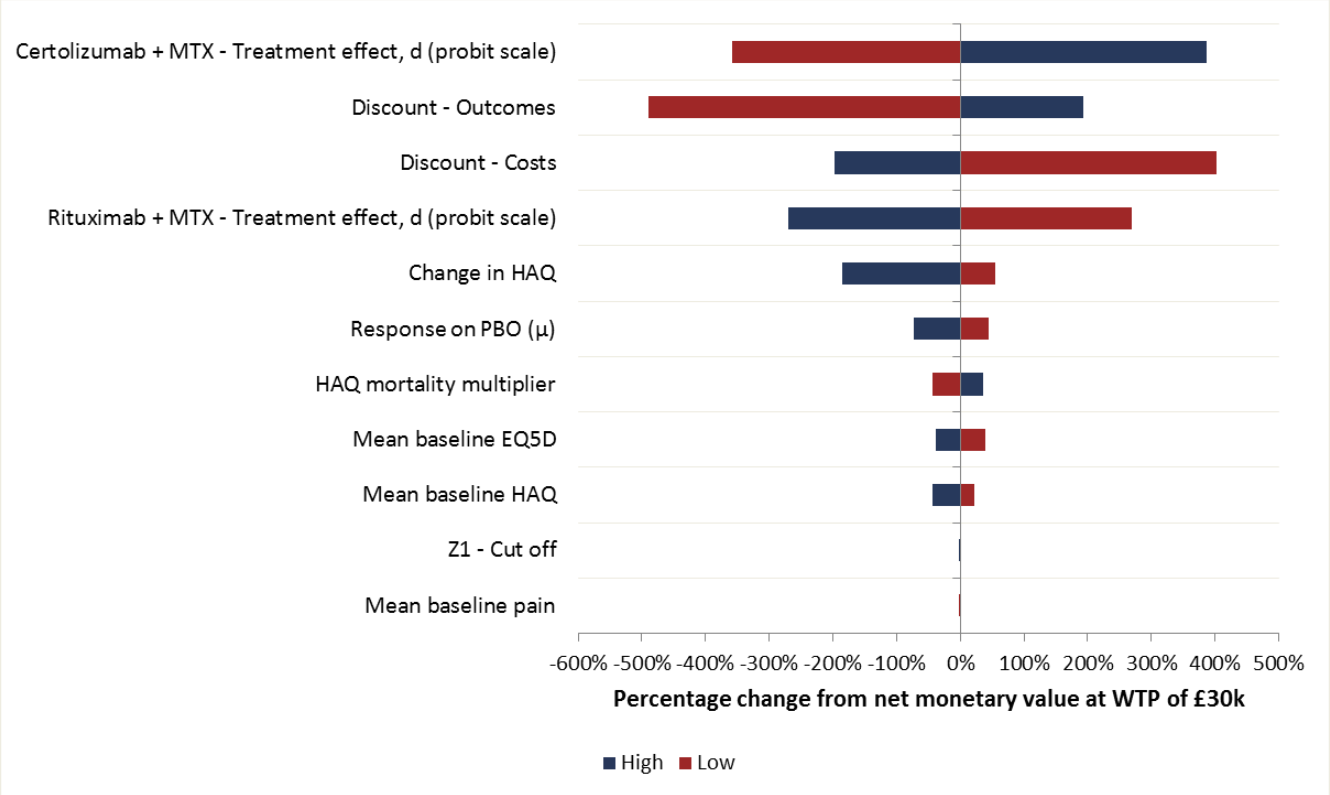


Figure 74: Tornado diagram showing the percentage change in net monetary benefit at a willingness to pay threshold of £30,000 per QALY gained, based on variation in individual model parameters – population B (CZP+MTX versus TOC+MTX)

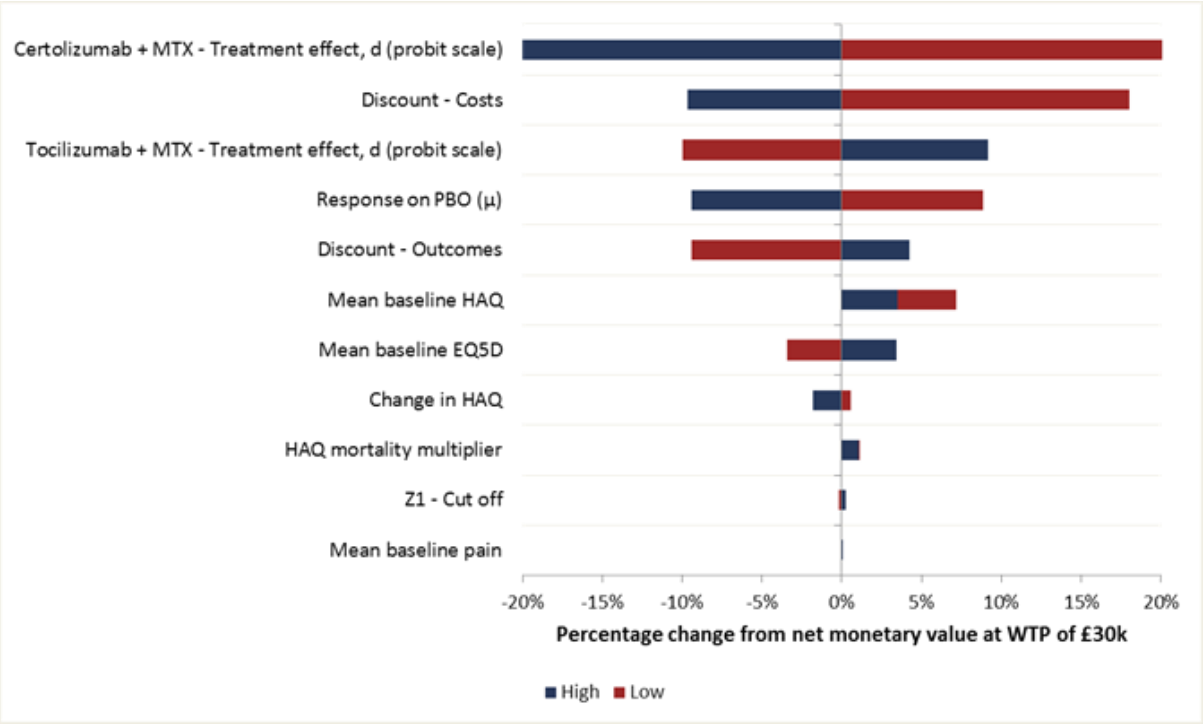
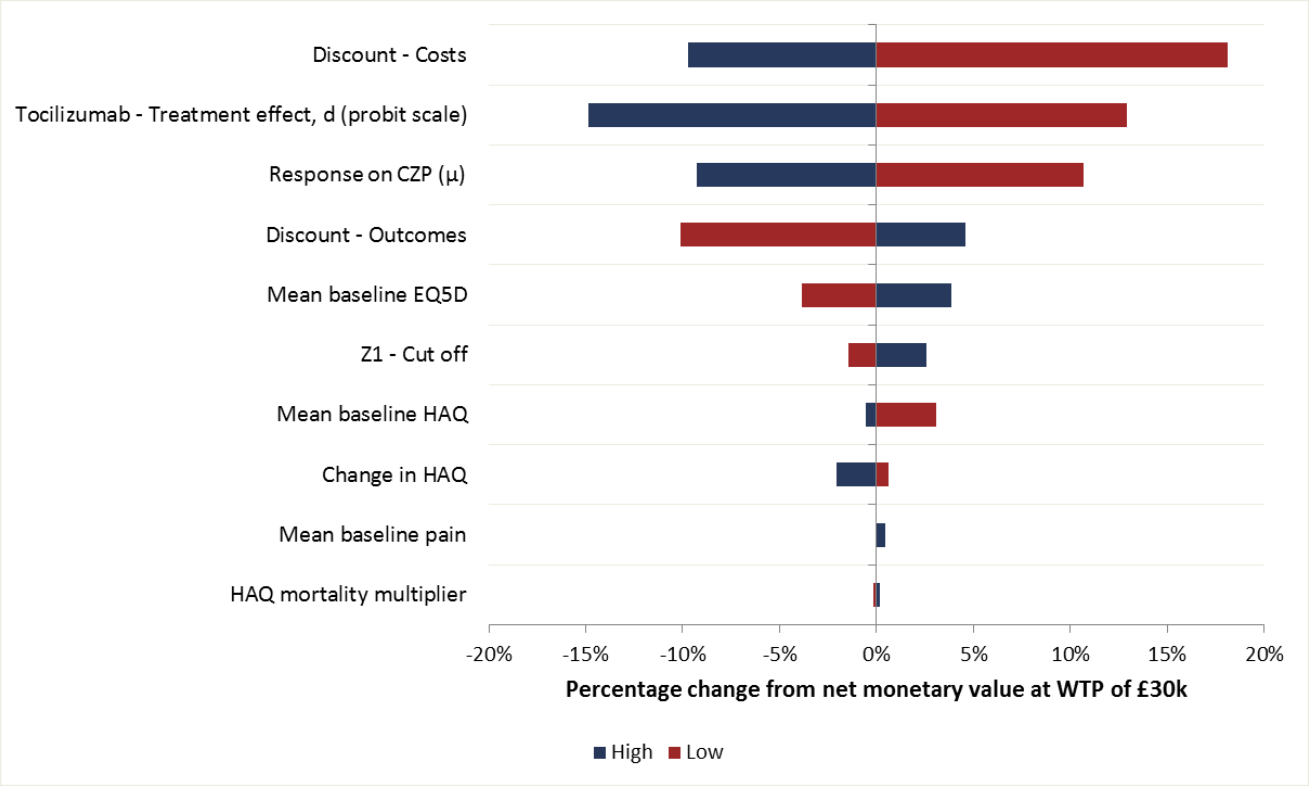


Figure 75: Tornado diagram showing the percentage change in net monetary benefit at a willingness to pay threshold of £30,000 per QALY gained, based on variation in individual model parameters – population C (CZP versus TOC)



5.8.3 Scenario analysis

A series of structural uncertainties were identified during the development of the model. These include uncertainties over the:

- The comparative efficacy of CZP versus PBO at six months (assume a class effect with other TNFi therapies)
- Modelling EQ-5D utility for response via HAQ score, as opposed to EQ-5D from PREDICT
- Utility gain achieved by week 6 of initial therapy
- Mapping from HAQ to EQ-5D (the use of alternative algorithms)
- The effect of cDMARDs and palliative care on HAQ score
- Duration of therapy after discontinuation of initial response (Are there differences in the long-term duration of therapy between TNFi and non-TNFis?)
- The frequency of re-treatment with RTX (varied from six to nine months re-treatment frequency)
- The use of subcutaneous formulations of TOC and ABA

- The use of biosimilars to IFX
- Accounting for the PASs for TOC and ABA

In addition, scenario analyses were performed around the following variables

- Time horizon (5-years and 10-years)
- Discount rates (combinations of 1.5% and 6.0% for costs and QALYs)
- Perspective (societal)

The results of the scenario analysis are provided in Table 101, and presented as fully incremental ICERs.

The results of the scenario analysis are summarised as follows:

- **All populations**: The results of the economic analysis were sensitive to assumptions on the model time horizon, perspective (societal versus NHS), and duration of bDMARD after response
- **Population A:** The results of the economic analysis for population A were also sensitive to assumptions on discount rates (ICER: £19k to £62k), choice of perspective (societal ICER: £4,729), mapping algorithm (ICER: up to £242k), and on the efficacy of CZP (ICER=£170k when assuming CZP is of equivalent efficacy to GOL)
- **Population B and C:** The results of the economic analysis for populations B and C were generally robust to the different scenarios considered in the scenario analysis, with CZP remaining cost effective at conventional thresholds of £20k to £30k

Table 101. Scenario analysis results – all populations

See next page

Parameter	Base case estimate	Sensitivity estimate	Population A results (ICER CZP vs. comparator)	Population B results (Fully incremental analysis)							Population C results (Fully incremental analysis)			
			Base case	CZP + MTX	ABA + MTX	ADA + MTX	ETA + MTX	GOL + MTX	IFX + MTX	TOC + MTX	CZP	ADA	ETA	TOC
			£34,516	CE at WTP of £3k to £129k	Dominated	CE at WTP < £3k	Dominated	CE at WTP < £3k	Dominated	CE at WTP > £129k	CE at WTP of £5k to £123k	CE at WTP < £5k	Dominated	CE at WTP > £123k
Time horizon	Lifetime	5 years	CZP dominates	CE at WTP <£80k	Dominated	Dominated	Dominated	Dominated	Dominated	CE at WTP > £80k	CE at WTP of £1k to £180k	CE at WTP < £1k	Dominated	CE at WTP > £180k
		10 years	£51,108	CE at WTP of £5k to £164	Dominated	CE at WTP < £5k	Dominated	CE at WTP < £5k	Dominated	CE at WTP > £164k	CE at WTP of £7k to £159k	CE at WTP < £7k	Dominated	CE at WTP > £159k
Discount rate	Costs and QALYs 3.5%	Costs 1.5% and QALYs 1.5%	£32,351	CE at WTP of £4k to £120	Dominated	CE at WTP < £4k	Dominated	CE at WTP < £4k	Dominated	CE at WTP > £120k	CE at WTP of £5k to £114k	CE at WTP < £5k	Dominated	CE at WTP > £114k
		Costs 1.5% and QALYs 6%	£61,915	CE at WTP of £5k to £161	Dominated	CE at WTP < £5k	Dominated	CE at WTP < £5k	Dominated	CE at WTP > £161k	CE at WTP of £7k to £155k	CE at WTP < £7k	Dominated	CE at WTP > £155k
		Costs 6% and QALYs 1.5%	£18,882	CE at WTP of £3k to £103	Dominated	CE at WTP < £3k	Dominated	CE at WTP < £3k	Dominated	CE at WTP > £103k	CE at WTP of £3k to £99k	CE at WTP < £3k	Dominated	CE at WTP > £99k
		Costs 6% and QALYs 6%	£36,137	CE at WTP of £4k to £139	Dominated	CE at WTP < £4k	Dominated	CE at WTP < £4k	Dominated	CE at WTP > £139k	CE at WTP of £3k to £99k	CE at WTP < £3k	Dominated	CE at WTP > £99k
Perspective	NHS/PSS	Societal	£4,729	CE at WTP < £118k	Dominated	Dominated	Dominated	Dominated	Dominated	CE at WTP > £118k	CE at WTP of £5k to £134k	CE at WTP < £5k	Dominated	CE at WTP > £135k

EQ-5D via HAQ	Direct – PREDICT	HAQ score from REALISTIC mapped to EQ-5D	£33,199	CE at WTP of £6k to £204	Dominated	CE at WTP < £6k	Dominated	CE at WTP < £6k	Dominated	CE at WTP > £204k	CE at WTP of £8k to £189k	CE at WTP < £8k	Dominated	CE at WTP > £189k
Utility gain at 6 weeks	100%	25%	£34,430	CE at WTP of £3k to £132	Dominated	CE at WTP < £3k	Dominated	CE at WTP < £3k	Dominated	CE at WTP > £132k	CE at WTP of £5k to £126k	CE at WTP < £5k	Dominated	CE at WTP > £126k
Mapping from HAQ to EQ-5D	Brennan et al	Hernandez class 1	£67,061	CE at WTP of £4k to £144	Dominated	CE at WTP < £4k	Dominated	CE at WTP < £4k	Dominated	CE at WTP > £144k	CE at WTP of £6k to £140k	CE at WTP < £6k	Dominated	CE at WTP > £140k
		Hernandez class 2	£242,348	CE at WTP of £5k to £158	Dominated	CE at WTP < £5k	Dominated	CE at WTP < £5k	Dominated	CE at WTP > £158k	CE at WTP of £7k to £155k	CE at WTP < £7k	Dominated	CE at WTP > £155k
		Hernandez class 3	£45,151	CE at WTP of £4k to £136	Dominated	CE at WTP < £4k	Dominated	CE at WTP < £4k	Dominated	CE at WTP > £136k	CE at WTP of £5k to £131k	CE at WTP < £5k	Dominated	CE at WTP > £131k
		Hernandez class 4	£36,462	CE at WTP of £4k to £131	Dominated	CE at WTP < £4k	Dominated	CE at WTP < £4k	Dominated	CE at WTP > £131k	CE at WTP of £5k to £1251k	CE at WTP < £5k	Dominated	CE at WTP > £125k
Efficacy of CZP	Based on mapped data from REALISTIC	For populations A and B, assume same effect for CZP + MTX as GOL + MTX (class effect with other TNFis) For population C, assume ADA and ETA to be of equivalent efficacy to CZP mono	£169,690	CE at WTP <£62k	Extendedly dominated by CZP and TOC	Dominated	Dominated	Dominated	Dominated	CE at WTP >£62k	CE at WTP < £793k	Dominated	Dominated	CE at WTP > £793k

	Based on mapped data from REALISTIC	Using NMA with J-RAPID study (probit effect of -1.131 for CZP vs. PBO)	£29,613	CE at WTP of £4k to £182	Dominated	CE at WTP < £4k	Dominated	CE at WTP < £4k	Dominated	CE at WTP > £182k	CE at WTP > £7k	CE at WTP < £7k	Dominated	Dominated
Duration of non-TNF therapy longer than TNF (scale parameter, Weibull)	0.4416	0.4416 (TNFi) 0.2208 (non-TNFi biologics)	CZP is dominated	CE at WTP of £4k to £43k	Extendedly dominated by CZP and TOC	CE at WTP < £4k	Dominated	CE at WTP < £4k	Dominated	CE at WTP > £43k	CE at WTP of £5k to £44k	CE at WTP < £5k	Dominated	CE at WTP > £44k
Duration of TNF therapy longer than non-TNF (scale parameter, Weibull)	0.4416	0.3003 (TNFi) 0.4416 (non TNFi biologics)	£19,673	CE at WTP of £7k to £2M	Dominated	CE at WTP < £7k	Dominated	CE at WTP < £7k	Dominated	CE at WTP > £2M	CE at WTP > £7k	Dominated	Dominated	Dominated
Costs associated with unused vials	Unused vials are lost (ie. waste)	Unused vials are consumed elsewhere in the health system (ie. no waste)	£34,110	CE at WTP of £4k to £98k	Dominated	CE at WTP < £4k	Dominated	CE at WTP < £4k	Dominated	CE at WTP > £98k	CE at WTP of £5k to £94k	CE at WTP < £5k	Dominated	CE at WTP > £94k
RTX treatment frequency	Every 6 months	Every 9 months	£49,618	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
TOC / ABA (SC)	IV	SC	NA	CE at WTP of £4k to £69k	Dominated	CE at WTP < £4k	Dominated	CE at WTP < £4k	Dominated	CE at WTP > £69k	CE at WTP of £5k to £225k	CE at WTP < £5k	Dominated	CE at WTP > £225k
TOC and ABA PAS (IV)	List price	17% discount	NA	CE at WTP of £4k to £88k	Dominated	CE at WTP < £4k	Dominated	CE at WTP < £4k	Dominated	CE at WTP > £88k	CE at WTP of £5k to £85k	CE at WTP < £5k	Dominated	CE at WTP > £85k
Biosimilar to IFX	IFX + MTX	Biosimilar + MTX	NA	CE at WTP of £3k to £129	Dominated	CE at WTP < £3k	Dominated	CE at WTP < £3k	Dominated	CE at WTP > £129k	NA	NA	NA	NA
HAQ progression on cDMARDs	0.045 per annum	0 per annum	£53,578	CE at WTP of £5k to £140k	Dominated	CE at WTP < £5k	Dominated	CE at WTP < £3k	Dominated	CE at WTP > £140k	CE at WTP of £5k to £133k	CE at WTP < £5k	Dominated	CE at WTP > £133k

HAQ progression on palliative care	0.06 per annum	0	£57,156	CE at WTP of £7k to £155k	Dominated	CE at WTP < £7k	Dominated	CE at WTP < £7k	Dominated	CE at WTP > £155k	CE at WTP of £10k to £155k	CE at WTP < £10k	Dominated	CE at WTP > £155k
Maximum mean HAQ score in the population	2.76	3.0	£34,183	CE at WTP of £4k to £130k	Dominated	CE at WTP < £4k	Dominated	CE at WTP < £4k	Dominated	CE at WTP > £130k	CE at WTP of £5k to £123k	CE at WTP < £5k	Dominated	CE at WTP > £123k

5.8.4 Summary of sensitivity analyses results

Across all populations, the results of the PSA were consistent with the results of the deterministic analyses. For population A, CZP had a low probability (3.0%) of being cost effective at the £20,000 per QALY threshold but the figure increased to 37% for a threshold of £30,000 per QALY. For populations B and C, CZP was very close to 100% probability of being cost effective at thresholds of both £20,000 and £30,000 per QALY.

The DSA showed that the model was most sensitive to changes in the comparative efficacy of CZP and comparator treatments, the discounting rates applied to costs and outcomes, and the change in EQ-5D associated with a unit change in HAQ.

The scenario analysis results show that the cost effectiveness model is sensitive in all three populations to:

- A short time horizon
- A societal perspective
- Hernandez class 2 mapping algorithm
- Treatment effect of CZP set equal to GOL (Population B and C)
- Duration of therapy after response

In the analyses for populations B and C the model was also sensitive when EQ-5D were based on mapped HAQ scores from the REALISTIC trial as well as when the admin route for ABA and TOC were changed from IV to SC.

Unlike models from previous submissions (TA375 and TA195), the results for population A were not heavily impacted by setting re-treatment of RTX to every 9 months.

5.9 *Subgroup analysis*

Not applicable.

5.10 *Validation of de novo cost effectiveness analysis*

The design of the economic model has been informed by a review of previous published economic models with a UK perspective (Kielhorn, 2008; Emery, 2009) and three recent NICE appraisals in RA (MTA195, TA247, and TA225). There are no previous appraisals or publications relating to CZP in the population in the scope of this appraisal and for this reason it is not possible to compare the results of this analysis directly with previous work.

Some previous appraisals in RA have presented an individual patient simulation model (IPS), but the more common approach is a cohort-based Markov state transition model. We have preferred this approach on the basis of transparency and parsimony. We are not aware that IPS has a particular benefit over the chosen approach.

Clinical pathways and the sequence of subsequent treatments were validated with a clinical expert practicing in England.

5.11 *Interpretation and conclusions of economic evidence*

In general, the results of previous economic studies are similar, but limited. In particular, RTX was found to either dominate or be considered cost effective compared to other bDMARDs. In most studies, there was limited information on whether the studied populations had contraindications or intolerance to either RTX or MTX, and it was therefore challenging to relate the studies to the populations in the scope of the appraisal.

Due to differences in study design (ie. comparators, time horizon), it was not feasible to compare ICERs between studies. In general, the conclusions of the evaluations are consistent with existing NICE guidance, which recommends the following treatment options for patients who have not responded to a previous TNFi; RTX, ABA, ADA, ETA, GOL, TOC, and IFX.

The analysis is relevant to all of the patient groups eligible for CZP.

The analysis is reflective of clinical practice in England and is consistent with current NICE guidance on the treatment of adult patients with RA who have had an inadequate response to a TNFi. All of the resource use assumptions and unit prices are current to the NHS in England.

Strengths

- De novo analysis designed to address the decision problem defined in the final scope of the appraisal
- Model structure and inputs were informed by a systematic literature review of published economic evaluations and cost and resource use studies
- The EULAR response to first therapy was estimated from an NMA of trials identified in a systematic literature review.
- Baseline health-state utilities and the function used to map utility to response was derived directly from analysis of patient-level data from the PREDICT study. Beyond the period of the trial, mapping algorithms were identified from a recent systematic review (Pennington)

- The economic model includes extensive sensitivity analysis and scenario analysis

Weaknesses:

- No directly comparable previous evaluations against which to validate results
- Mapping from HAQ to EQ-5D in the post-trial period is sensitive to the choice of mapping algorithm. There are many possible algorithms available (addressed in SA)
- There is conflicting evidence about the relationship between therapy and rates of discontinuation. In the base case, these rates are assumed to be the same for all treatment (tested in SA)

The base case analysis has been subject to extensive sensitivity analysis, and there is limited opportunity to conduct additional analysis given the data which are currently available.

6 Assessment of factors relevant to the NHS and other parties

Executive summary

- The net budget savings of recommending CZP in all population groups (A, B, and C) in the appraisal ranged from approximately £470,000 in 2016 to £2.76 million in 2020, translated into a cumulative net budgetary saving of £8,37 million over 5 years.
- In population A, the net budget impact ranged from -£32,000 (2016) to +£382,000 (2020).
- In the combined population B (withdrawn from RTX due to adverse events) and population C (eligible for monotherapy biologic), treatment with CZP is expected to yield annual cost-savings of between £438,000 (2016) and £3.14 million (2020), when compared with current practice (no CZP use). The expected cumulative cost-savings from CZP use in these populations is approximately £9.10 million over 5 years.

6.1 *Introduction*

This section contains an analysis of factors relevant to the NHS and other parties that fall outside the remit of clinical and cost effectiveness. This comprises an assessment of the potential budget impact of recommending CZP as a treatment for TNFi-IR. To our knowledge, there are no organisational, societal (ie. equity) or ethical issues relating to the use of CZP in this population.

6.2 *Model structure*

A budget impact model was developed to evaluate the net financial impact of recommending CZP as a treatment option for patients who have had an inadequate response to their first TNFi. In line with the scope of the appraisal, the budget impact evaluation considers the use of CZP in three populations:

- Population A: adults previously treated with other DMARDs, including at least one TNFi, who are eligible for RTX treatment
- Population B: adults for whom RTX has been withdrawn due to AEs
- Population C: adults for whom RTX therapy cannot be given because MTX is contraindicated or withdrawn

Following earlier sections of the submission, the budget impact model and corresponding evaluation does not consider the use of CZP in patients who have not adequately responded to bDMARDs, including RTX (population D).

The outcomes of the budget impact analysis are presented from the perspective of the National Health Service and Personal Social Services in England and Wales. Results are presented over a 5-year time horizon, representing the potential uptake of CZP in TNFi-IR between 2016 and 2020.

The structure of the budget impact model is largely based on the NICE costing template that was developed for TA195. The original costing template was expanded to include comparators relevant to this appraisal (ie. CZP, RTX, and GOL), and to include budget estimates for years 2 to 5.

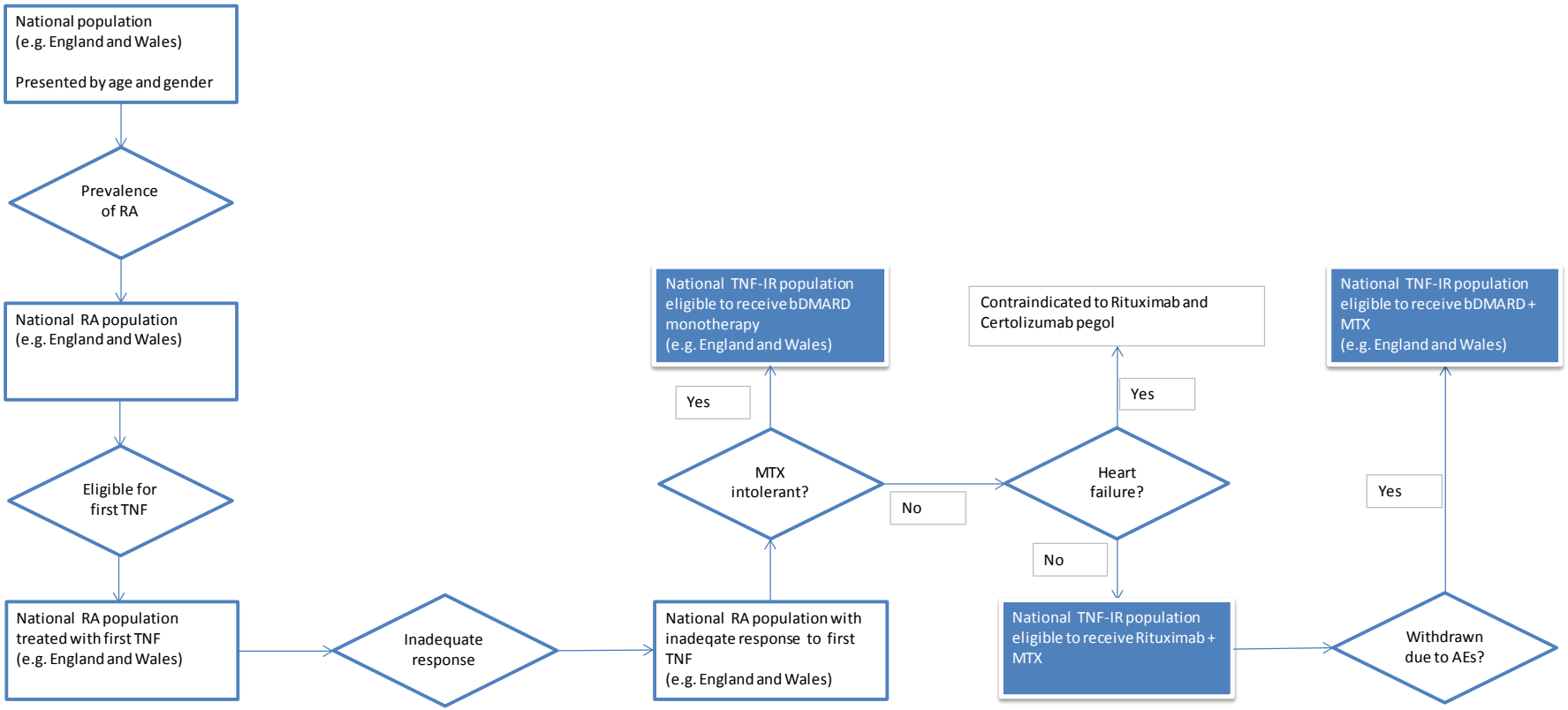
An illustration of the structure of the budget impact model is shown in Figure 76. Further detail on the calculation steps and parameter estimates used in the base case evaluation are presented in the following sections.

A summary of the key assumptions in the budget impact evaluation are presented in Table 102.

Table 102: Model fixed assumptions

Assumption	Justification / sensitivity analysis
All epidemiological data are expected to remain constant over time (ie. percentage prevalence of RA)	Simplifying assumption
The number of patients with diagnosed RA is expected to increase in line with the expected growth in the number of adults in England and Wales, over the same period	Simplifying assumption
All treatments are administered as per current NICE guidance, and there is no off-label or out of scope use of therapies (ie. no use of RTX monotherapy, or treatments in combination with leflunomide)	In line with the scope of the appraisal
All patients are assumed to receive a full year's course of treatment in each budget year	Simplifying assumption that is applied to all treatments in the evaluation
RTX re-treatment assumed to occur at a frequency of once every 6 months, resulting in two courses of treatment in a given budget year	A sensitivity analysis was performed where RTX costs were rescaled based on retreatment frequency of once every 9 months
PASs for TOC and ABA based on 17% discount in list price (sensitivity analysis)	Based on assumption that discount offered by manufacturers was to offset the costs of intravenous administration
Recommendations on the use of CZP are not expected to impact NHS expenditure in treating adverse events or the monitoring of drug and disease activity	Consistent with assumptions taken in previous appraisals, assuming no difference in monitoring and adverse events between bDMARDs
There are no costs associated with the administration, and training of patients, using pre-filled subcutaneous injection pens	Based on clinical expert opinion

Figure 76: Budget impact model structure



6.2.1 Patient population eligible for treatment

The eligible population in the budget impact evaluation comprises all RA patients in England and Wales who have failed on a first TNFi, and are eligible for CZP as per its current marketing authorisation. The total numbers of patients in this population are derived through a series of calculations, as outlined in Appendix 8.15.1.1

A summary of the parameter estimates used to estimate the numbers of eligible patients is provided in Appendix 8.15.1.2.

The budget impact model predicts that the total number of patients with RA will increase from 405,656 (2016) to 417,135 (2020) over the next 5 years. The increase in prevalence is in line with the expected growth in the numbers of adults in England and Wales over the same period. Of the total RA population, it is estimated that between 12,170 (2016) and 12,514 (2020) patients would be TNFi-IR (Appendix 8.15.1.2).

Of the TNFi-IR population (n=12,170 in 2016), it is estimated that 51% of patients would be eligible for bDMARD monotherapy (n=6,207 in 2016) due to intolerance or contraindication to MTX (population C). Of the remaining 49% (n=5,963 in 2016), a proportion of patients (0.77%, n=94) would be contraindicated to certain biologics, including RTX and CZP (Appendix 8.15.1.2). This patient group is excluded from the model and the corresponding budget impact calculations.

The remaining 48% of patients (n=5,869 in 2016) comprise all patients currently eligible for RTX and MTX (population A). Of this patient group, it is estimated that 12.0% (n=704 in 2016) will withdraw from RTX because of adverse events (Appendix 8.15.1.2). According to existing NICE guidance, these patients are eligible for treatment with bDMARDs in combination with MTX (population B).

A summary of the size of the total eligible population is provided in Table 103.

Table 103: total number of eligible patients by population in the appraisal

Total eligible population (excluding patients with contraindications)		**Year 1**	**Year 2**	**Year 3**	**Year 4**	**Year 5**
		12,780**	12,870	12,960	13,051	13,142
Population A	Adults previously treated with other DMARDs including at least one TNF	5,869	5,911	5,952	5,994	6,036
Population B	Adults for whom RTX is contraindicated or withdrawn due to AEs	704	709	714	719	724

Total eligible population		Year 1	Year 2	Year 3	Year 4	Year 5
Population C	Adults for whom RTX therapy cannot be given because MTX is contraindicated or withdrawn	6,207	6,250	6,294	6,338	6,382

** 12,780 = 12,170 – 94 + 704 (see text above). The patients in Population B are also included in the number of patients in Population A, so these patients are double counted in the total eligible population.

The net budget impact of recommending CZP in TNFi-IR is calculated by comparing the total budget for current market (ie. no recommendation on the use of CZP), versus a new market where CZP is recommended and used in TNFi-IR.

In each scenario, the total budget costs are calculated using estimates of the patient population eligible for treatment and data on the market share and associated costs of therapy. Further detail on the market share and cost parameters is provided in sections 6.2.2 and 6.2.3.

6.2.2 Market share assumptions

The market share projections for current market were derived from UK prescription sales data, which reports the market share of bDMARDs in all patients with RA. These data were used to model market shares in TNFi-IR, on the assumption that market shares are largely consistent between first TNF and TNFi-IR uses.

Market share data were not available by specific population groups. As only population A can receive RTX and these patients cannot receive other bDMARDs, it is assumed that all RTX use was attributed to population A. The market share data for all other comparators, however, is based on patients eligible for treatment in populations B or C, and therefore budget impact results in these two groups are presented together. The market share projections for population A and population B/C are presented in Table 104 and Table 105. In both groups, market share projections are presented for current clinical practice and for a scenario where CZP is recommended and used in practice.

For the new market, it is assumed that CZP would represent up to [redacted] of current market share in patients with RA after 5 years. In all populations, the market share for CZP was assumed to increase gradually over the next 5 years. CZP was assumed to displace all existing therapies at rates proportional to their current market shares.

The market share forecasts for population A are presented in Table 104.

Table 104: Market share projections for biologic treatment in Population A: people eligible for RTX plus MTX

Treatment	2016	2017	2018	2019	2020
Market share in current market					
CZP + MTX	0.0%	0.0%	0.0%	0.0%	0.0%
RTX + MTX	100.0%	100.0%	100.0%	100.0%	100.0%
Market share in new market					
CZP + MTX	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
RTX + MTX	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

The market share forecasts for population B/C are presented in Table 105. IFX and biosimilars to IFX, Remsima and Inflectra, were also considered as comparators for this population; however, there were no reliable data to population their market share so they have been excluded from the analysis.

In the new market, it is predicted that the market share for CZP would increase gradually from [redacted] in 2016 to [redacted] in 2020. The corresponding market shares for alternative therapies are assumed to decrease at rates proportional to their current uptake.

Table 105: Market share projections for biologic treatment in Population B/C: withdrawn from RTX treatment due to AEs or contraindicated to/AEs with MTX

Treatment	2016	2017	2018	2019	20120
Market share in current clinical practice					
CZP	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
TOC	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
ADA	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
ETA	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
ABA	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
GOL	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Market share where CZP is recommended and used in practice					
CZP	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
TOC	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
ADA	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
ETA	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
ABA	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
GOL	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

6.2.3 Resource utilisation and unit costs

The budget impact model considers the costs associated with the acquisition and administration of drug therapy. Other costs such as the costs of adverse event management or routine monitoring were not considered in the model.

A summary of the dosing schedule and unit costs of drug therapy is shown in Table 106.[22]

The unit costs of drug therapy were obtained from the latest British National Formulary. The drug dosing schedules were obtained from the summary of product characteristics for each treatment in evaluation.

For drugs that are dosed based on weight (eg. intravenous ABA, intravenous TOC, IFX), the average dose was derived based on a mean weight of 83.7 kg (based on baseline data from the REALISTIC study). The acquisition costs of these drugs were calculated assuming no vial wastage (ie. leftover drug from open vials were used to treat other patients).

For drugs that have irregular administration or for which dosing varies, the following assumptions were made:

- A course of RTX consists of two 1000 mg intravenous infusions given 2 weeks apart. Repeat courses of RTX may be provided if residual disease activity remains or disease activity returns. In the base case, it was assumed that repeat courses of RTX occurred every 6 months (ie. two courses per budget year). A sensitivity analysis was performed where RTX re-treatment occurred every 9 months (based on expert opinion).
- ADA may be administered weekly or every two weeks (fortnightly); the model assumes fortnightly administration.
- ETA may be administered as 25 mg twice weekly or 50 mg once weekly; the model assumes once weekly administration although total treatment costs will not be affected by the frequency of administration.
- MTX is administered initially as 7.5 mg once weekly, but dosage is adjusted according to response with a maximum weekly dose of 20 mg. The model assumes a dosage of 15 mg.

ABA and TOC are available in both intravenous and subcutaneous formulations. Due to differences in dosing and administration schedules, the costs of the intravenous formulations may differ from the costs of subcutaneous formulations (Table 106). In the base case, the costs of ABA and TOC were based on infusion-administration. A sensitivity analysis was performed assuming both therapies are administered via their subcutaneous formulation.

The dosing schedules for CZP and ABA include an initial loading dose that leads to differences in the number of doses administered in the first and subsequent years of treatment. The financial impact of different dosing between years of treatment was incorporated in the evaluation by calculating the numbers of newly and previously treated patients in each budget year of the evaluation.

In the first budget year, it was assumed that all patients were newly treated. For all subsequent years, the numbers of previously treated patients were estimated by multiplying the number of patients treated in the previous year with the annual probability of continuing therapy from one budget year to the next. The numbers of newly treated patients were derived from the total number treated (estimated from market share data) minus the number previously treated.

The annual probability of continuing treatment from one budget year to the next was estimated at 64.3% based on data from the BSRBR. The same probability was applied to all treatments in the budget impact calculation.

Current recommendations on the use of CZP, GOL, TOC, and ABA in England and Wales are subject to PASs that include free stock arrangements or direct discounts to list prices. The PASs for TOC and ABA are simple discount schemes where the manufacturer provides drug at a discount to the list price. Currently, the discounts offered by the manufacturers are confidential. The PASs for CZP and GOL are free stock arrangements:

- CZP: 10 free vials of CZP given during the first 12 weeks of therapy
- GOL: manufacturer provides the 100 mg dose at the same price as the 50mg dose. The 100 mg dose is recommended for patients who weight more than 100 kg and fail to respond to three or four administrations of the 50mg dose.

For all evaluations (base case and sensitivity analyses), the costs of CZP and GOL are adjusted to take account of the PAS. The PASs for TOC and ABA were excluded from the base case analysis. A sensitivity analysis was performed where the costs of TOC and ABA were reduced to capture the impact of discounts on the results of the budget impact evaluation. The actual discount offered by the manufacturers of TOC and ABA is confidential, and therefore an assumed discount of 17% was applied (equivalent to the cost of drug administration).

Table 106: Drug dosing schedule and costs

Intervention	Route of administration	Dosing schedule	Dose (mg per administration)	Number of administrations per year	Pack size/formulation	Cost per pack/formulation
Biological DMARDs						
CZP	Subcutaneous injection	400 mg at Weeks 0, 2 and 4, and 200 mg every two weeks thereafter	200	Newly treated: 29.0 Previously treated: 26.0	200 mg pre-filled syringe	£357.50
RTX	Intravenous infusion	The model assumed RTX is administered once every 6 months	2000 (two 1000-mg doses separated by 2 weeks)	2.0	500 mg vial	£873.15
TOC	Intravenous infusion	Administered every 4 weeks	Body weight dependent: 8 mg/kg	13.0	80 mg vial	£102.40
	Subcutaneous injection	Administered every week	162	52.0	162mg pre-filled syringe	£228.28
ADA	Subcutaneous injection	Two different frequencies of administration (weekly or Q2W) are possible; the model assumes Q2W administration.	40	26.0	40 mg pre-filled syringe	£352.14
ETA	Subcutaneous injection	Administered as 25 mg twice weekly or 50 mg once weekly; the model assumes once weekly administration	50	52.0	25 mg pre-filled syringe	£89.38
ABA	Intravenous infusion	Administered every 4 weeks	Body weight dependent; 500 mg if <60 kg, 750 mg if 60-100 kg and 1 g if >100 kg. Administered at Weeks 0, 2, 4, and every 4 weeks thereafter	Newly treated: 14.0 Previously treated: 13.0	250 mg vial	£302.40
	Subcutaneous injection	Administered every week	125	52.0	125 mg pre-filled syringe	£302.40
GOL	Subcutaneous injection	Administered once a month, on the same date each month	50	12.0	50 mg pre-filled syringe	£762.92
CDMARDs						
MTX	Oral	Administered once weekly	7.5 mg once weekly, adjusted according to response; maximum weekly dose 20 mg. The model assumes a dosage of 15 mg.	52.0	2.5 mg tablets – 24-tab pack	£2.40

The budget impact model considered the cost of treatment administration. This comprised the costs associated with the administration of IV-based therapies (£174 per infusion, see section 5.5.2.2). It was assumed that treatments administered via subcutaneous (SC) injection (ie. CZP) are not associated with administration costs, as patients are expected to administer therapy at home or with the support of a home help scheme, which is funded by the pharmaceutical industry. Further, it was assumed that patients would not require nurse training for the administration of SC treatments given that the majority of patients would have received training for their prior TNFi. Any training required would likely be minimal or covered by routine follow-up care.

The total annual costs of drug acquisition and administration for treatments are shown in Table 107.

Table 107: Summary of annual drug acquisition and administration costs for each treatment comparator in the model

Treatment	Annual acquisition costs (£)	Annual administration costs (£)	Total annual costs (£)
CZP			
Newly treated patients (with PAS)	6,793	0	6,793
Previously treated patients	9,295	0	9,295
RTX	6,985	347	7,332
TOC – IV (without PAS)	11,155	2,002	13,412
TOC - SC (without PAS)	11,871	0	11,871
ADA	9,156	0	9,156
ETA	9,296	0	9,296
ABA - IV (without PAS)			
Newly treated patients	12,701	2,156	15,131
Previously treated patients	11,794	2,002	14,050
ABA – SC (without PAS)	15,725	0	15,725
GOL (with PAS)	9,156	0	9,156
MTX	31	0	31

Note: the costs for CZP account for the PAS agreed with the NHS; the costs of toclizumab and ABA are based on the publically available list prices as reported by the British National Formulary; therefore their reported costs do not take into account the confidential price discount PAS agreed between the manufacturers and the Department of Health.

6.3 *Results*

The results of the budget impact evaluation are presented by population and aggregated across all three populations in the appraisal.

For brevity, the drug and administration costs are aggregated in terms of CZP, alternative TNFi (ADA, ETA, GOL), non- TNFis (ABA, TOC, RTX) and biosimilars (inflectra, remsima).

Within each population, a series of sensitivity analyses were performed to assess the impact of assumptions on the conclusion of the budget impact evaluation. These analyses include:

- Re-treatment with RTX every 9 months
- Using subcutaneous formulation of ABA and TOC
- PASs for ABA and TOC

A summary of the budget impact results is provided in the following sections.

Summary of budget impact results

The net budget impact of recommending CZP as an alternative treatment option in TNFi-IR is presented in Table 108.

Table 108: Summary of budget impact calculations by population and budget year

Population	2016	2017	2018	2019	2020	Cumulative
	Net annual budget impact (budget of current market minus budget of new market)					
Population A	-£31,689	-£1,282	£124,665	£252,374	£381,866	£725,934
Population B/C	-£437,981	-£1,195,833	-£1,836,074	-£2,485,220	-£3,143,364	-£9,098,472
All populations	-£469,670	-£1,197,115	-£1,711,410	-£2,232,846	-£2,761,498	-£8,372,539

In population B/C (withdrawn from RTX due to AEs/eligible for monotherapy biologic), treatment with CZP is expected to yield annual cost-savings of £438,000 in 2016 rising to £3.14 million in 2020, when compared with current practice (no CZP use). Over the next five years, the expected cumulative cost-savings from CZP use in these populations is approximately £9.10 million. A number of sensitivity analyses were performed to assess the impact of the formulations (IV versus SC) and PASs for TOC and ABA on the results of the evaluation. In all sensitivity analyses, the scenario that included CZP remained cost-saving compared with current practice.

In population A (eligible for RTX), the net annual budget impact of recommending CZP as an alternative to RTX varied from approximately -£32,000 (2016) to +£381,000 (2020). When reducing the mean frequency of RTX treatment from 6 to 9

months, the net budget impact of CZP varied from approximately +£112,000 in 2016 to +£1.71 million in 2020. The additional costs were attributed to the acquisition of CZP.

The aggregated net budget impact of recommending CZP in all population groups (A and B/C) in the appraisal ranged from approximately -£470,000 in 2016 to -£2.76 million in 2020. The cost-savings associated with recommending CZP in population B/C exceeded the net costs expected from recommending CZP in population A.

A full breakdown on the numbers of patients treated, the total budget costs and net budget impact costs in each population of the appraisal is presented in the following sections of the document.

6.3.1 Population A

The numbers of patients treated with CZP and RTX are presented in Table 109.

Based on the market share forecasts, it is estimated that between [redacted] patients would be treated with CZP, if recommended as a treatment alternative to RTX.

Table 109: Numbers of patients treated in each budget year of the evaluation [Population A]

Treatment	2016	2017	2018	2019	2020
Number of patients treated in current market					
CZP	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
RTX	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Number of patients treated in a new market					
CZP	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
RTX	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

The net budget impact of recommending CZP as an alternative treatment to RTX in TNFi-IR is presented in Table 110. Between 2016 and 2020, the net annual budget impact is estimated to vary between -£31,689 (2016) and +£381,866 (2020). Over the next five years, the net cumulative budget impact is estimated at £725,934.

Table 110: Population A: Base case: breakdown net total budget costs for 2016 to 2020

	2016	2017	2018	2019	2020	Cumulative
Current market						
Total drug acquisition costs	£41,182,349	£41,470,625	£41,760,920	£42,053,246	£42,347,619	£208,814,758
CZP	£0	£0	£0	£0	£0	£0
Alternative TNFi	£0	£0	£0	£0	£0	£0
Non-TNFi	£41,182,349	£41,470,625	£41,760,920	£42,053,246	£42,347,619	£208,814,758
Total drug administration costs	£2,037,870	£2,052,135	£2,066,500	£2,080,966	£2,095,532	£10,333,003
CZP	£0	£0	£0	£0	£0	£0
Alternative TNFi	£0	£0	£0	£0	£0	£0
Non-TNFi	£2,037,870	£2,052,135	£2,066,500	£2,080,966	£2,095,532	£10,333,003
Total cost	**£43,220,219**	**£43,522,760**	**£43,827,420**	**£44,134,212**	**£44,443,151**	**£219,147,762**
New market						
Total drug acquisition costs	£41,171,038	£41,530,907	£41,988,909	£42,451,288	£42,918,083	£210,060,225
CZP	£400,513	£1,304,401	£2,316,035	£3,341,769	£4,381,749	£11,744,468
Alternative TNFi	£0	£0	£0	£0	£0	£0
Non-TNFi	£40,770,525	£40,226,506	£39,672,874	£39,109,519	£38,536,333	£198,315,757
Total drug administration costs	£2,017,491	£1,990,571	£1,963,175	£1,935,298	£1,906,934	£9,813,470
CZP	£0	£0	£0	£0	£0	£0
Alternative TNFi	£0	£0	£0	£0	£0	£0
Non-TNFi	£2,017,491	£1,990,571	£1,963,175	£1,935,298	£1,906,934	£9,813,470
Total cost	**£43,188,530**	**£43,521,478**	**£43,952,084**	**£44,386,586**	**£44,825,017**	**£219,873,695**
Net budget impact	**-£31,689**	**-£1,282**	**£124,665**	**£252,374**	**£381,866**	**£725,934**

Sensitivity analysis assuming re-treatment with RTX every 9 months

The frequency of re-treatment with RTX was reduced from every 6 to 9 months, leading to a reduction in the annual costs of RTX (£6,985 in the base case [assuming two doses] to £4,540 assuming an average of 1.3 treatments per year [ie. 12/9]).

The results of the budget impact analysis assuming fewer administrations of RTX are presented in Table 114.

Table 111: Population A: sensitivity analysis (re-treatment with RTX every 9 months) - Net budget impact for 2016 to 2020

	2016	2017	2018	2019	2020
Total cost –current market	£28,157,237	£28,354,337	£28,552,818	£28,752,687	£28,953,956
Total cost – with CZP recommendation	£28,276,177	£28,808,108	£29,441,212	£30,081,768	£30,729,850
Net budget impact	£118,941	£453,771	£888,395	£1,329,081	£1,775,893

With fewer administrations of RTX, the net budget impact of recommending CZP as a treatment alternative to RTX is estimated to vary from £118,941 in 2016 to £1,775,893 in 2020.

6.3.2 Population B/C

The numbers of patients treated with CZP, alternative TNFis, and non-TNFis, given in combination with MTX (population B) or as monotherapy (population C), are presented in Table 109.

Based on the market share forecasts, it is estimated that between [redacted] patients would be treated with CZP, if recommended as a treatment option in patients withdrawn from RTX.

Table 112: Numbers of patients treated in each budget year of the evaluation [Population B/C]

Treatment	2015	2016	2017	2018	2019
Market share in current market					
CZP	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Alternative TNFi	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Non-TNFi	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Market share in new market					
CZP	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Alternative TNFi	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Non-TNFi	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

The net budget impact of recommending CZP as an alternative treatment in TNFi-IR who have withdrawn from RTX is presented in Table 110. Between 2016 and 2020, the net annual budget impact is estimated to vary between -£437,981 (2016) and -£3,143,364 (2020), representing cost-savings to the NHS. Over the next five years, a positive recommendation for CZP is expected to yield net cost-savings of £9,098,472

Table 113: Population B/C: Basecase: breakdown net total budget costs and net budget impact for 2016 to 2020 [Population B/C]

	2016	2017	2018	2019	2020	Cumulative
Current market						
Total drug acquisition costs	£78,154,082	£77,349,891	£77,891,341	£78,436,580	£78,985,636	£390,817,530
CZP	£0	£0	£0	£0	£0	£0
Alternative TNFi	£13,111,115	£13,202,893	£13,295,313	£13,388,380	£13,482,099	£66,479,799
Non-TNFi	£65,042,967	£64,146,999	£64,596,028	£65,048,200	£65,503,537	£324,337,731
Total drug administration costs	£12,801,769	£12,632,805	£12,721,235	£12,810,284	£12,899,956	£63,866,049
CZP	£0	£0	£0	£0	£0	£0
Alternative TNFi	£0	£0	£0	£0	£0	£0
Non-TNFi	£12,801,769	£12,632,805	£12,721,235	£12,810,284	£12,899,956	£63,866,049
Total cost	**£90,955,851**	**£89,982,697**	**£90,612,576**	**£91,246,864**	**£91,885,592**	**£454,683,579**
New market						
Total drug acquisition costs	£77,844,118	£76,538,214	£76,696,536	£76,853,324	£77,008,549	£384,940,741
CZP	£471,577	£1,535,845	£2,726,977	£3,934,710	£5,159,217	£13,828,326
Alternative TNFi	£12,980,004	£12,806,806	£12,630,547	£12,451,193	£12,268,710	£63,137,260
Non-TNFi	£64,392,537	£62,195,564	£61,339,012	£60,467,421	£59,580,622	£307,975,156
Total drug administration costs	£12,673,751	£12,248,650	£12,079,966	£11,908,320	£11,733,679	£60,644,365
CZP	£0	£0	£0	£0	£0	£0
Alternative TNFi	£0	£0	£0	£0	£0	£0
Non-TNFi	£12,673,751	£12,248,650	£12,079,966	£11,908,320	£11,733,679	£60,644,365
Total cost	**£90,517,870**	**£88,786,864**	**£88,776,502**	**£88,761,644**	**£88,742,228**	**£445,585,106**
Net budget impact	-£437,981	-£1,195,833	-£1,836,074	-£2,485,220	-£3,143,364	-£9,098,472

Sensitivity analysis using subcutaneous formulation of ABA and TOC

The annual drug acquisition costs for the SC formulations of ABA and TOC were estimated at £15,725 (52 pre-filled syringes at £302.40 per syringe) and £11,871 (52 pre-filled syringes at £228.28 per syringe), respectively. In line with other SC drugs, the costs of drug administration were set to £0. The total costs (drug and administration) of intravenous ABA and TOC were £15,131 (first year) and £13,412, respectively.

The results of the budget impact analysis assuming SC formulations for ABA and TOC are presented in Table 114.

Table 114: Population B/C: sensitivity analysis (SC formulations for ABA and TOC) - Net budget impact for 2016 to 2020

	2016	2017	2018	2019	2020
Total cost – current practice	£87,431,318	£88,043,337	£88,659,640	£89,280,258	£89,905,220
Total cost – with CZP recommendation	£87,028,582	£86,937,882	£86,953,635	£86,965,349	£86,972,967
Net budget impact	-£402,736	-£1,105,455	-£1,706,005	-£2,314,908	-£2,932,253

A positive recommendation on the use of CZP in TNFi-IR who have withdrawn from RTX is expected to yield annual cost-savings of between £402,736 (2016) and £2,932,253 (2020).

Sensitivity analysis (Patient access schemes for ABA and TOC)

To evaluate the impact of the TOC and ABA PASs on the results of the budget impact analysis, the drug acquisitions costs of these therapies were subject to a 17% discount.

The discount was set at 17% on the assumption that the PASs for these drugs were designed to offset the additional costs of intravenous administration (~£2000 per year). The actual discounts offered by the manufacturers of TOC and ABA are confidential.

With the 17% price discount, the total costs of therapy were reduced from £15,131 to £12,972 (first year) in the case of ABA, and from £13,412 to £11,516 in the case of TOC.

The results of the budget impact analysis assuming price discounts for ABA and TOC are presented in Table 115.

Table 115: Population B/C: sensitivity analysis (patient access scheme for ABA and TOC) - Net budget impact for 2016 to 2020

	2016	2017	2018	2019	2020

	2016	2017	2018	2019	2020
Total cost – current practice	£79,927,688	£79,107,053	£79,660,802	£80,218,428	£80,779,957
Total cost – with CZP recommendation	£79,599,989	£78,242,084	£78,376,943	£78,509,857	£78,640,792
Net budget impact	-£327,700	-£864,969	-£1,283,859	-£1,708,571	-£2,139,165

The new market is expected to yield annual cost-savings of between £327,700 (2016) and £2,139,165 (2020).

6.3.3 Limitations within the budget impact analysis.

A limitation of the budgetary analysis is the uncertainty surrounding the estimates of the current market shares of alternative biologic therapies and the assumed uptake of CZP in the different populations in scope following a positive recommendation from NICE. Furthermore, the patient access schemes for comparators were not publically available and could not have been accounted for in the analysis. The results were also sensitive to assumptions about the re-treatment frequency of rituximab.

7 References

1. Deighton C, Hyrich K, Ding T, et al. BSR and BHPR rheumatoid arthritis guidelines on eligibility criteria for the first biological therapy. Rheumatology (Oxford). 2010;49(6):1197-1199.
2. Bukhari M, Abernethy R, Deighton C, et al. BSR and BHPR guidelines on the use of rituximab in rheumatoid arthritis. Rheumatology (Oxford). 2011;50(12):2311-3.
3. UCB. CIMZIA Summary of Product Characteristics 2014.
4. Weinblatt ME, Fleischmann R, Huizinga TW, et al. Efficacy and safety of certolizumab pegol in a broad population of patients with active rheumatoid arthritis: results from the REALISTIC phase IIIb study. Rheumatology (Oxford). 2012.
5. Furst DE, Shaikh SA, Greenwald M, et al. Two dosing regimens of certolizumab pegol in patients with active rheumatoid arthritis. Arthritis Care Res (Hoboken). 2015;67(2):151-60.
6. Curtis JR, Churchill M, Kivitz A, et al. A Randomized Trial Comparing Disease Activity Measures to Assess and Predict Response in Rheumatoid Arthritis Patients Initiating Certolizumab Pegol. Arthritis Rheumatol. 2015.
7. Schiff MH, von Kempis J, Goldblum R, et al. Rheumatoid arthritis secondary non-responders to TNF can attain an efficacious and safe response by switching to certolizumab pegol: a phase IV, randomised, multicentre, double-blind, 12-week study, followed by a 12-week open-label phase. Ann Rheum Dis. 2014;73(12):2174-7.
8. Yamamoto K, Takeuchi T, Yamanaka H, et al. Efficacy and safety of certolizumab pegol plus methotrexate in Japanese rheumatoid arthritis patients with an inadequate response to methotrexate: the J-RAPID randomized, placebo-controlled trial. Modern Rheumatology. 2014;24(5):715-724.
9. Yamamoto K, Takeuchi T, Yamanaka H, et al. Efficacy and safety of certolizumab pegol without methotrexate co-administration in Japanese patients with active rheumatoid arthritis: the HIKARI randomized, placebo-controlled trial. Modern Rheumatology. 2014;24(4):552-560.
10. Bykerk VP, Cush J, Winthrop K, et al. Update on the safety profile of certolizumab pegol in rheumatoid arthritis: an integrated analysis from clinical trials. Ann Rheum Dis. 2015;74(1):96-103.
11. Simard JF, Arkema EV, Sundstrom A, et al. Ten years with biologics: to whom do data on effectiveness and safety apply? Rheumatology (Oxford). 2011;50(1):204-13.
12. Yun H, Xie F, Delzell E, et al. Comparative Risk of Hospitalized Infection Associated With Biologic Agents in Rheumatoid Arthritis Patients Enrolled in Medicare. Arthritis Rheumatol. 2016;68(1):56-66.
13. UCB. Cimzia® [Online] Accessed: 19th November 2015 http://www.ucb.com/investors/UCB-today/Cimzia.
14. NICE. TA375 Adalimumab, etanercept, infliximab, certolizumab pegol, golimumab, tocilizumab and abatacept for rheumatoid arthritis not previously treated with DMARDs or after conventional DMARDs only have failed. 2016.
15. Aletaha D,Funovits J,Keystone EC and Smolen JS. Disease activity early in the course of treatment predicts response to therapy after one year in rheumatoid arthritis patients. Arthritis Rheum. 2007;56(10):3226-35.

16. Kavanaugh A, Smolen J, Emery P, et al. Long-Term Benefits Over More Than 4 Years of Certolizumab Pegol Combination Therapy on Workplace and Household Productivity, and Participation in Social Activities in Rheumatoid Arthritis: Results from the Open-Label Extension Study of Rapid1. Value in Health. 2013;16(7):A570.
17. Cooper NJ. Economic burden of rheumatoid arthritis: a systematic review. Rheumatology (Oxford). 2000;39(1):28-33.
18. EMA. EPAR CIMZIA. http://www.ema.europa.eu/ema/index.jsp?curl=pages/medicines/human/medicines/001037/human_med_001294.jsp&murl=menus/medicines/medicines.jsp&mid=WC0b01ac058001d125. 2011.
19. NICE. TA186 Rheumatoid arthritis - certolizumab pegol: guidance, 2010.
20. NICE. TA383 TNF-alpha inhibitors for ankylosing spondylitis and non-radiographic axial spondyloarthritis. 2016.
21. NICE. ID579 Certolizumab pegol and secukinumab for treating active psoriatic arthritis following inadequate response to disease modifying antirheumatic drugs (final scope). 2015.
22. BNF. British National Formulary. 2016.
23. Goel N and Stephens S. Certolizumab pegol. MAbs. 2010;2(2):137-47.
24. Parton T,King L,Parker G and Nesbitt A. The PEG moiety of Certolizumab Pegol is rapidly cleared from the blood of humans by the kidneys once it is cleaved from the Fab'. EULAR 2009. 2009.
25. Palframan R, Airey M, Moore A, et al. Use of biofluorescence imaging to compare the distribution of certolizumab pegol, adalimumab, and infliximab in the inflamed paws of mice with collagen-induced arthritis. J Immunol Methods. 2009;348(1-2):36-41.
26. Carron P, Lambert B, De Vos F, et al. Scintigraphic Detection Of TNFα With a Radiolabeled Anti-TNFα In Patients With Active Peripheral Spondyloarthritis and Rheumatoid Arthritis. Abstract 1113. Arthritis Rheum. 2013;65(Suppl 10):S472.
27. Stephens S,Browns D,Nesbitt A and Foulkes R. Lack of Placental Transfer and Accumulation in Milk of an Anti-TNF PEGylated Fab' Fragment in Rats. European Crohn's and Colitis Organisation (ECCO) Annual Meeting, March 1-3, 2007. Innsbruck, Austria. Abstract P152.
28. Porter C, Kopotsha T, Smith B, et al. No Significant Transfer of Certolizumab Pegol Compared with IgG in the Perfused Human Placenta In Vitro. Poster THU0191. Ann Rheum Dis. 2010;69(Suppl10):S674.
29. Wolf D and Mahadevan U. Certolizumab pegol use in pregnancy: low levels detected in cord blood. Abstract 718. Arthritis Rheum. 2010;62(Suppl 10):S299.
30. Mahadevan U, Wolf DC, Dubinsky M, et al. Placental transfer of anti-tumor necrosis factor agents in pregnant patients with inflammatory bowel disease. Clin Gastroenterol Hepatol. 2013;11(3):286-92; quiz e24.
31. Clowse ME, Wolf DC, Forger F, et al. Pregnancy Outcomes in Subjects Exposed to Certolizumab Pegol. J Rheumatol. 2015.
32. Red Dot Award: Communication Design. Design Innovations 2009 Competition, Design Zentrum Nordrhein Westfalen, Germany. Cimzia Packaging [Online] Accessed: 5th November 2015 http://www.oxouk.com/Ourawards.aspx. [Accessed online at on

33. Red Dot Award: Product Design. Design Innovations 2010 Competition, Design Zentrum Nordrhein Westfalen, Germany. Cimzia Syringe [Online] Accessed: 5th November 2015 http://www.oxouk.com/Ourawards.aspx. [Accessed online at on
34. National Audit Office. Services for people with rheumatoid arthritis. 2009.
35. NICE. Rheumatoid arthritis: the management of rheumatoid arthritis in adults http://www.nice.org.uk/CG79, 2009.
36. Symmons D, Turner G, Webb R, et al. The prevalence of rheumatoid arthritis in the United Kingdom: new estimates for a new century. Rheumatology (Oxford). 2002;41(7):793-800.
37. Huscher D, Merkesdal S, Thiele K, et al. Cost of illness in rheumatoid arthritis, ankylosing spondylitis, psoriatic arthritis and systemic lupus erythematosus in Germany. Ann Rheum Dis. 2006;65(9):1175-83.
38. Guillemin F, Durieux S, Daures JP, et al. Costs of rheumatoid arthritis in France: a multicenter study of 1109 patients managed by hospital-based rheumatologists. J Rheumatol. 2004;31(7):1297-304.
39. Hallert E,Husberg M and Skogh T. Costs and course of disease and function in early rheumatoid arthritis: a 3-year follow-up (the Swedish TIRA project). Rheumatology (Oxford). 2006;45(3):325-31.
40. NRAS response to DSU and Assessment Reports for NICE MTA ID537. 2015:1083-1114.
41. Seymour HE,Worsley A,Smith JM and Thomas SH. Anti-TNF agents for rheumatoid arthritis. Br J Clin Pharmacol. 2001;51(3):201-8.
42. Hyrich KL, Lunt M, Watson KD, et al. Outcomes after switching from one anti-tumor necrosis factor alpha agent to a second anti-tumor necrosis factor alpha agent in patients with rheumatoid arthritis: results from a large UK national cohort study. Arthritis Rheum. 2007;56(1):13-20.
43. NICE. TA195 Rheumatoid arthritis - adalimumab, etanercept, infliximab, rituximab and abatacept; guidance. 2010.
44. NICE. TA225 Golimumab for the treatment of rheumatoid arthritis after the failure of previous disease-modifying anti-rheumatic drugs. 2011.
45. NICE. TA247 Tocilizumab for the treatment of rheumatoid arthritis. 2012.
46. Smolen JS, Landewe R, Breedveld FC, et al. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2013 update. Ann Rheum Dis. 2014;73(3):492-509.
47. Khurana R and Berney SM. Clinical aspects of rheumatoid arthritis. Pathophysiology. 2005;12(3):153-65.
48. Cojocaru M, Cojocaru IM, Silosi I, et al. Extra-articular Manifestations in Rheumatoid Arthritis. Maedica (Buchar). 2010;5(4):286-91.
49. National Rheumatoid Arthritis Society. The economic burden of rheumatoid arthritis. 2010.
50. Lundkvist J,Kastang F and Kobelt G. The burden of rheumatoid arthritis and access to treatment: health burden and costs. Eur J Health Econ. 2008;8 Suppl 2:S49-60.
51. Luqmani R, Hennell S, Estrach C, et al. British Society for Rheumatology and British Health Professionals in Rheumatology guideline for the management of rheumatoid arthritis (after the first 2 years). Rheumatology (Oxford). 2009;48(4):436-9.

52. Rindfleisch JA and Muller D. Diagnosis and management of rheumatoid arthritis. Am Fam Physician. 2005;72(6):1037-47.
53. van Vollenhoven R,Harju A,Brannemark S and Klareskog L. Treatment with infliximab (Remicade) when etanercept (Enbrel) has failed or vice versa: data from the STURE registry showing that switching tumour necrosis factor alpha blockers can make sense. Ann Rheum Dis. 2003;62(12):1195-8.
54. Curtis JR and Singh JA. Use of biologics in rheumatoid arthritis: current and emerging paradigms of care. Clin Ther. 2011;33(6):679-707.
55. Merck. Merck Manuals. Rheumatoid Arthritis (RA). 2008.
56. Dickens C,McGowan L,Clark-Carter D and Creed F. Depression in rheumatoid arthritis: a systematic review of the literature with meta-analysis. Psychosom Med. 2002;64(1):52-60.
57. Li X,Gignac MA and Anis AH. The indirect costs of arthritis resulting from unemployment, reduced performance, and occupational changes while at work. Med Care. 2006;44(4):304-10.
58. Bansback N, Zhang W, Walsh D, et al. Factors associated with absenteeism, presenteeism and activity impairment in patients in the first years of RA. Rheumatology (Oxford). 2012;51(2):375-84.
59. National Rheumatoid Arthritis Society Survey. Employment and Rheumatoid Arthritis: A National Employment Picture. 2007.
60. Hetland ML, Christensen IJ, Tarp U, et al. Direct comparison of treatment responses, remission rates, and drug adherence in patients with rheumatoid arthritis treated with adalimumab, etanercept, or infliximab: results from eight years of surveillance of clinical practice in the nationwide Danish DANBIO registry. Arthritis Rheum. 2010;62(1):22-32.
61. Rendas-Baum R, Wallenstein GV, Koncz T, et al. Evaluating the efficacy of sequential biologic therapies for rheumatoid arthritis patients with an inadequate response to tumor necrosis factor-alpha inhibitors. Arthritis Res Ther. 2011;13(1):R25.
62. Furst DE, Gaylis N, Bray V, et al. Open-label, pilot protocol of patients with rheumatoid arthritis who switch to infliximab after an incomplete response to etanercept: the opposite study. Ann Rheum Dis. 2007;66(7):893-9.
63. NICE. NICE RA Commissioning Algorithm. 2013.
64. Weinblatt ME, Fleischmann R, van Vollenhoven RF, et al. Twenty-eight-week results from the REALISTIC phase IIIb randomized trial: efficacy, safety and predictability of response to certolizumab pegol in a diverse rheumatoid arthritis population. Arthritis Res Ther. 2015;17(1):325.
65. Combe B, Dougados M, Logeart I, et al. AB0516 An open-label randomized controlled study to evaluate the efficacy of etanercept (ETN) versus rituximab (RTX), in patients with active rheumatoid arthritis previously treated with RTX and TNF blockers. Annals of the Rheumatic Diseases. 2013;71(Suppl 3):667.
66. Gottenberg J, Brocq O, Perdriger A, et al. n the Multicenter Randomized Controlled Rotation or Change Trial, a Non-TNF Targeted Therapy Has a Higher Efficacy Than a Second Anti-TNF at 3, 6 and 12 Months [abstract]. Arthritis Rheumatol. 2015;67(suppl 10).
67. Chatzidionysiou K, Kristensen LE, Eriksson J, et al. Effectiveness and survival-on-drug of certolizumab pegol in rheumatoid arthritis in clinical practice: results from the national Swedish register. Scand J Rheumatol. 2015;44(6):431-7.

68. Pope J. Rapid reductions in fatigue and sleep problems, and correlation with improvements in patient-reported outcomes in patients with active rheumatoid arthritis (RA) treated with certolizumab pegol in the REALISTIC 12-week phase IIIb randomised controlled study. . Ann Rheum Dis 2011;70(Suppl. 3):258.
69. Tanaka Y, Yamamoto K, Takeuchi T, et al. Long-term efficacy and safety of certolizumab pegol in Japanese rheumatoid arthritis patients with an inadequate response to methotrexate: 52-week results from an open-label extension of the J-RAPID study. Mod Rheumatol. 2014;24(5):734-43.
70. Takeuchi T, Yamamoto K, Yamanaka H, et al. Post-hoc analysis showing better clinical response with the loading dose of certolizumab pegol in Japanese patients with active rheumatoid arthritis. Modern Rheumatology. 2015:1-8.
71. Takeuchi T, Yamamoto K, Yamanaka H, et al. Clinical Response At 12 Weeks Predicts Long-Term Remission and the Extent of Radiographic Progression in Japanese Patients with Rheumatoid Arthritis Treated with Certolizumab Pegol with and without Methotrexate Coadministration, In ARTHRITIS AND RHEUMATISM, WILEY-BLACKWELL 111 RIVER ST, HOBOKEN 07030-5774, NJ USA, 2012.
72. Tanaka Y, Yamamoto K, Takeuchi T, et al. Long-term efficacy and safety of certolizumab pegol in Japanese rheumatoid arthritis patients who could not receive methotrexate: 52-week results from an open-label extension of the HIKARI study. Mod Rheumatol. 2014;24(5):725-33.
73. UCB. Statistical analysis plan (SAP), Study: C87094, Product: certolizumab pegol. 2010.
74. Yamanaka H, Yamamoto K, Takeuchi T, et al. Improved Physical Function, Pain, and Health Related Quality of Life with Certolizumab Pegol in Japanese Rheumatoid Arthritis Patients with an Inadequate Response to Methotrexate: Results from the JRAPID Study. Ann Rheum Dis. 2012;71(Suppl3):664.
75. Yamanaka H, Yamamoto K, Takeuchi T, et al. Improved Physical Function, Pain, and Health Related Quality of Life With Certolizumab Pegol in Japanese Rheumatoid Arthritis Patients without Methotrexate Co-Administration: Results from the HIKARI Study. Ann Rheum Dis. 2012;71(Suppl3):664.
76. Pope J, Bingham CO, 3rd, Fleischmann RM, et al. Impact of certolizumab pegol on patient-reported outcomes in rheumatoid arthritis and correlation with clinical measures of disease activity. Arthritis Res Ther. 2015;17:343.
77. Kavanaugh A, Mease P, Strand V, et al. Effect Of Certolizumab Pegol On Workplace And Household Productivity In United States Patients With Rheumatoid Arthritis With Or Without Prior Anti-Tnf Exposure: Results From The Predict Study. Value in Health. 2014;17(3):A53.
78. DerSimonian R and Laird N. Meta-analysis in clinical trials revisited. Contemp Clin Trials. 2015;45(Pt A):139-45.
79. Higgins J and Green S. Cochrane handbook for systematic reviews of interventions. The Cochrane collaboration. Version 5.1.0. <http://www.cochrane-handbook.org/> Accessed Feb 2016. 2011.
80. Centre for Reviews and Dissemination. CRD's guidance for undertaking reviews in health care. 2009.
81. Jadad AR, Moore RA, Carroll D, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? Control Clin Trials. 1996;17(1):1-12.

82. Dias S,Welton NJ,Sutton AJ and Ades A. NICE DSU technical support document 2: a generalised linear modelling framework for pairwise and network meta-analysis of randomised controlled trials. URL http://www.nicedsu.org.uk. 2011.
83. Bucher HC,Guyatt GH,Griffith LE and Walter SD. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. J Clin Epidemiol. 1997;50(6):683-91.
84. Jansen J, Fleurence R, Devine B, et al. Interpreting indirect treatment comparisons and network meta-analysis for health-care decision making: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 1. Value Health. 2011;14(4):417-428.
85. Singh JA, Wells GA, Christensen R, et al. Adverse effects of biologics: a network meta-analysis and Cochrane overview. Cochrane Database Syst Rev. 2012(2):CD008794.
86. Weinblatt M, Fleischmann R, van Vollenhoven R, et al. Certolizumab pegol as monotherapy or with concomitant DMARDs in patients with active rheumatoid arthritis with or without prior TNF inhibitor use: Analyses of the REALISTIC 12-week phase IIIb randomised controlled study. FRI0214. Presented at EULAR 2011. 2011.
87. Curtis JR, Koetse W, Tambiah J, et al. Prediction of Week 52 Treatment Response Based on a Week 12 Assessment in Rheumatoid Arthritis Patients Receiving Certolizumab Pegol: Comparison of a Patient-Reported Instrument Versus Physician-Based Disease Activity Assessment. Poster 431, In American College of Rheumatology, San Diego, CA, USA, 2013.
88. Mariette X, Vencovsky J, Lortholary O, et al. The incidence of tuberculosis in patients treated with certolizumab pegol across indications: impact of baseline skin test results, more stringent screening criteria and geographic region. RMD Open. 2015;1(1):e000044.
89. Curtis JR,de Longueville M,O'Brien C and Haraoui B. Improvement in Disease Activity and the Long-Term Risk of Serious Infectious Events in Rheumatoid Arthritis Patients Treated with Certolizumab Pegol [Abstract 561]. Arthritis Rheumatol. 2015;67(Suppl 10).
90. Malottki K, Barton P, Tsourapas A, et al. Adalimumab, etanercept, infliximab, rituximab and abatacept for the treatment of rheumatoid arthritis after the failure of a tumour necrosis factor inhibitor: a systematic review and economic evaluation. Health Technol Assess. 2011;15(14):1-278.
91. Emery P, Keystone E, Tony HP, et al. IL-6 receptor inhibition with tocilizumab improves treatment outcomes in patients with rheumatoid arthritis refractory to anti-tumour necrosis factor biologicals: results from a 24-week multicentre randomised placebo-controlled trial. Ann Rheum Dis. 2008;67(11):1516-23.
92. NICE. NICE pathways - Drug treatment for RA. 2016. 2-11-2015. Ref Type: Online Source.
93. Hyrich K,Symmons D,Watson K and Silman A. Baseline comorbidity levels in biologic and standard DMARD treated patients with rheumatoid arthritis: results from a national patient register. Ann Rheum Dis. 2006;65(7):895-8.
94. Hurst NP, Kind P, Ruta D, et al. Measuring health-related quality of life in rheumatoid arthritis: validity, responsiveness and reliability of EuroQol (EQ-5D). Br J Rheumatol. 1997;36(5):551-9.

95. Hernandez Alava M,Wailoo A and Ara R. Tails from the Peak District: Adjusted Limited Dependent Variable Mixture Models of EQ-5D Health State Utility Values. Value in Health. 2010;15(3):550-561.
96. Kobelt G,Lindgren P,Singh A and Klareskog L. Cost effectiveness of etanercept (Enbrel) in combination with methotrexate in the treatment of active rheumatoid arthritis based on the TEMPO trial. . Ann Rheum Dis. 2005;64(8):1174-9.
97. Barbieri M,Wong J and Drummond M. The cost-effectiveness of infliximab for severe treatment-resistant rheumatoid arthritis in the UK. Pharmacoeconomics. 2005;23(6):607-18.
98. Kobelt G,Jonsson L,Young A and Eberhardt K. The cost-effectiveness of infliximab (Remicade) in the treatment of rheumatoid arthritis in Sweden and the United Kingdom based on the ATTRACT study. Rheumatology (Oxford). 2003;42(2):326-35.
99. Wong J,Singh G and Kavanaugh A. Estimating the cost-effectiveness of 54 weeks of infliximab for rheumatoid arthritis. Am J Med 2002;113(5):400-8.
100. Smolen JS, Kay J, Matteson EL, et al. Insights into the efficacy of golimumab plus methotrexate in patients with active rheumatoid arthritis who discontinued prior anti-tumour necrosis factor therapy: post-hoc analyses from the GO-AFTER study. Ann Rheum Dis. 2014;73(10):1811-8.
101. Boers M. The time has come to limit the placebo period in rheumatoid arthritis trials to 3 months: a systematic comparison of 3- and 6-month response rates in trials of biological agents. Ann Rheum Dis. 2010;69(1):186-92.
102. Genovese MC, Schiff M, Luggen M, et al. Longterm safety and efficacy of abatacept through 5 years of treatment in patients with rheumatoid arthritis and an inadequate response to tumor necrosis factor inhibitor therapy. J Rheumatol. 2012;39(8):1546-54.
103. Keystone EC, Cohen SB, Emery P, et al. Multiple courses of rituximab produce sustained clinical and radiographic efficacy and safety in patients with rheumatoid arthritis and an inadequate response to 1 or more tumor necrosis factor inhibitors: 5-year data from the REFLEX study. J Rheumatol. 2012;39(12):2238-46.
104. Ramiro S, Landewe R, van der Heijde D, et al. Discontinuation rates of biologics in patients with rheumatoid arthritis: are TNF inhibitors different from non-TNF inhibitors? RMD Open. 2015;1(1):e000155.
105. Wailoo A and Tosh J. DECISION SUPPORT UNIT. 2008.
106. Hernandez Alava M,Wailoo A,Wolfe F and Michaud K. A comparison of direct and indirect methods for the estimation of health utilities from clinical outcomes. Med Decis Making. 2014;34(7):919-30.
107. Edwards CJ, Arden NK, Fisher D, et al. The changing use of disease-modifying anti-rheumatic drugs in individuals with rheumatoid arthritis from the United Kingdom General Practice Research Database. Rheumatology (Oxford). 2005;44(11):1394-8.
108. Office for National Statistics. Population Estimates for UK, England and Wales, Scotland and Northern Ireland, Mid-2014. 25-6-2015. 3-12-2015. Online Source. 2014.
109. Wolfe F,Michaud K,Gefeller O and Choi HK. Predicting mortality in patients with rheumatoid arthritis. Arthritis Rheum. 2003;48(6):1530-42.

110. Sokka T,Hakkinen A,Krishnan E and Hannonen P. Similar prediction of mortality by the health assessment questionnaire in patients with rheumatoid arthritis and the general population. Ann Rheum Dis. 2004;63(5):494-7.
111. Norton S, Sacker A, Dixey J, et al. Trajectories of functional limitation in early rheumatoid arthritis and their association with mortality. Rheumatology (Oxford). 2013;52(11):2016-24.
112. Pennington B and Davis S. Mapping from the Health Assessment Questionnaire to the EQ-5D: the impact of different algorithms on cost-effectiveness results. Value Health. 2014;17(8):762-71.
113. Kind P,Dolan P,Gudex C and Williams A. Variations in population health status: results from a United Kingdom national questionnaire survey. Bmj. 1998;316(7133):736-41.
114. Brennan A, Bansback N, Nixon R, et al. Modelling the cost effectiveness of TNF-alpha antagonists in the management of rheumatoid arthritis: results from the British Society for Rheumatology Biologics Registry. Rheumatology (Oxford). 2007;46(8):1345-54.
115. Hernandez Alava M,Wailoo A,Wolfe F and Michaud K. The relationship between EQ-5D, HAQ and pain in patients with rheumatoid arthritis. Rheumatology (Oxford). 2013;52(5):944-50.
116. Kobelt G, Jonsson L, Lindgren P, et al. Modeling the progression of rheumatoid arthritis: a two-country model to estimate costs and consequences of rheumatoid arthritis. Arthritis Rheum. 2002;46(9):2310-9.
117. Wiles N,Cooper N and Symmons D. Resource use within the Norfolk Arthritis Register (NOAR) Cohort during the first five years of disease. Report for Roche. 2005.
118. Chakravarty K, McDonald H, Pullar T, et al. BSR/BHPR guideline for disease-modifying anti-rheumatic drug (DMARD) therapy in consultation with the British Association of Dermatologists. Rheumatology (Oxford). 2008;47(6):924-5.
119. Office for National Statistics. Consumer Price Index. 9-2-2016. Online Source.
120. Downs SH and Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. J Epidemiol Community Health. 1998;52(6):377-84.

8 appendices

8.1 ***CZP regulatory approval worldwide***

8.2 ***Search strategy for relevant (RCT) studies***

8.2.1 Embase® and MEDLINE® database (results taken from a search on 16th November 2015 of Embase.com)

8.2.2 Cochrane database (results taken from a search on 16th November 2015 of onlinelibrary.wiley.com/cochranelibrary)

8.2.3 MEDLINE® In-process (results taken from a search on 16th November 2015 of pubmed.com)

8.3 ***Further Information on Study Selection for the Clinical systematic literature review***

8.3.1 Study designs captured by the systematic literature review

8.4 ***Studies captured by the literature review***

8.4.1 List of included studies

8.4.2 List of Studies Excluded from the Indirect Analysis

8.5 ***Quality Assessment of RCTs***

8.5.1 Full Quality Assessments of all Eligible RCTs

8.6 ***Additional study methodology, statistical analysis and definitions***

8.6.1 Blinding schedule for REALISTIC

8.6.2 mACR assessment, as used in PREDICT

8.7 ***REALISTIC study***

8.7.1 Treatment response by monotherapy and MTX combination subgroups

8.7.1.1 REALISTIC study: ACR response rates by overall study population monotherapy and combination with MTX subgroups (NRI)

8.7.1.2 REALISTIC study: ACR response rates by TNFi experienced monotherapy and combination with MTX subgroups (NRI)

8.7.1.3 REALISTIC study: EULAR response rates by TNFi experienced monotherapy and combination with MTX subgroups (LOCF)

8.7.1.4 REALISTIC study: DAS28 scores by TNFi experienced monotherapy and combination with MTX subgroups (LOCF)

8.7.1.5 REALISTIC study: DAS28(ESR) remission by TNFi experienced monotherapy and combination with MTX subgroups (NRI)

8.7.1.6 REALISTIC study: CDAI by TNFi experienced monotherapy and combination with MTX subgroups (LOCF)

8.7.1.7 REALISTIC study: Patient's Assessment of Fatigue by TNFi experienced monotherapy and combination with MTX subgroups (LOCF)

8.7.1.8 REALISTIC study: Sleep Problem Index II score by TNFi experienced monotherapy and MTX subgroups (LOCF)

8.7.1.9 REALISTIC study: HAQ-DI by TNFi experienced monotherapy and combination with MTX subgroups (LOCF)

8.7.2 ACR response rates (LOCF)

8.12 *Methods, results, outcomes and quality assessment of the relevant trials in the indirect treatment comparison*

8.12.1 Summary of studies included in the indirect treatment comparison

8.12.1.1 Summary of patient inclusion/exclusion criteria in RCTs relevant to the indirect analyses

8.12.1.2 Summary of study design and baseline characteristics of patients in RCTs relevant to the indirect analyses (TNFi-IR population; current treatment in combination with MTX)

8.12.1.3 Quality Assessment of RCTs Included in the ITC

8.12.2 Overview of outcomes reported in the studies included in ITC

8.12.2.1 Overview of outcomes reported at 3 months in the studies included in ITC

8.12.2.2 Overview of outcomes reported at 6-months in the studies included in ITC

8.12.3 Network diagrams for the ITC (base case analyses)

8.12.3.1 Network diagram for ACR20 response at 3 months

8.12.3.2 Network diagram for ACR50 response at 3 months

8.12.3.3 Network diagram for EULAR (good/moderate) response at 3 months

8.12.3.4 Network diagram for EULAR (good) response at 3 months

8.12.3.5 Network diagram for ACR20 response at 6-months

8.12.3.6 Network diagram for ACR50 response at 6 months

8.12.3.7 Network diagram for ACR70 response at 6 months

8.12.3.8 Network diagram for EULAR (good/moderate) response at 6 months

8.12.3.9 Network diagram for EULAR (good) response at 6 months

8.12.4 Network diagrams for the ITC (sensitivity analyses)

8.12.4.1 Network diagram for EULAR (good/moderate) at 3-months (sensitivity analysis)

8.12.4.2 Network diagram for EULAR (good) response at 3 months (sensitivity analysis)

8.12.4.3 Network diagram for EULAR (good/moderate) response at 6-months (sensitivity analysis)

8.12.4.4 Network diagram for EULAR (good) response at 6 months (sensitivity analysis)

8.12.5 Model fit characteristics from NMA

8.13 *Programming language used in the analysis*

8.13.1 Model (fixed-effects)

8.13.2 Model (Random-effects)

8.13.3 Adjusted ITC

8.13.4 Derivation of missing data used for calculating dichotomous outcomes

8.14 *Cost effectiveness appendices*

8.14.1 Economic systematic literature review

8.14.1.1 Search strategy for economic evidence in Embase® and Medline® as on 3rd November 2015

8.14.1.2 Search strategy for economic evidence in Cochrane library as on 3rd November 2015

8.14.1.3 Economic evaluation search flow diagramSummary list of published economic evaluations

8.14.1.4 Summary list of published economic evaluations

8.14.1.5 Summary list of relevant economic evaluations submitted as part of previous NICE appraisals

8.14.1.6 Studies excluded from the MTC due to prior bDMARDs

8.14.1.7 Excluded studies and reasons for exclusion

8.14.1.8 Drummond quality assessment checklist

8.14.2 Clinical parameters and variables

8.14.2.1 Studies identified in the systematic review and considered in the NMA for the economic model

8.14.2.2 Mapping matrices

8.14.2.2.1. Mapping matrix for EULAR response at 3-months versus EULAR-response at 6-months for PBO (N=326; in RAPID 1 and 2 pooled data set)

8.14.2.2.2. Response probabilities for patients in REALISTIC (Week 28/six months)

8.14.2.3 Summary of EULAR data considered in the NMA

8.14.2.4 NMA Winbugs code

8.14.2.5 Sensitivity analysis NMA: Treatment effect for EULAR response at six months, and reported on the probit scale for comparisons of CZP versus PBO and CZP versus RTX, with and without the inclusion of the J-RAPID study

8.14.2.6 HAQ and pain score on response to first therapy

8.14.2.6.1. Change in HAQ by response

8.14.2.7 Mapping algorithms

8.14.2.8 Health-related quality of life data

8.14.2.8.1. Distribution of residuals for final regression model, presented alongside standard normal distribution for data

8.14.2.8.2. Observed by predicted change since baseline EQ-5D

8.14.2.9 List of health states and associated costs in the economic model

8.14.3 Questions from clinical expert interview

8.14.4 Cost-effectiveness model results

8.14.4.1 Summary of model results compared with clinical data – population A – undiscounted unless otherwise stated – per patient

8.14.4.2 Summary of model results compared with clinical data – population B – undiscounted unless otherwise stated – per patient

8.14.4.3 Summary of model results compared with clinical data – population C – per patient

8.15 *Budget impact model appendices*

8.15.1 Model structure

8.15.1.1 Calculation of total number of patients

8.15.1.2 Parameter estimates used to derive the number of patients in England and Wales eligible for CZP as a second TNFi

8.15.1.3 Summary of model results compared with clinical data – population A – undiscounted unless otherwise stated – per patient

8.15.1.4 Summary of model results compared with clinical data – population B – undiscounted unless otherwise stated – per patient

8.15.1.5 Summary of model results compared with clinical data – population C – per patient

8.16 *References for appendices*

NICE National Institute for Health and Care Excellence

Level 1A
City Tower
Manchester
M1 4BT
United Kingdom

+44 (0)300 323 0140

Single technology appraisal

Certolizumab pegol for treating rheumatoid arthritis after inadequate response to a TNF inhibitor [ID824]

Dear Vincent,

The Evidence Review Group, School of Health & Related Research Sheffield (ScHARR), and the technical team at NICE have looked at the submission received on 4 March 2016 from UCB. In general they felt that it is well presented and clear. However, the ERG and the NICE technical team would like further clarification on the clinical and cost effectiveness data (see questions listed at end of letter).

The ERG and the technical team at NICE will be addressing these issues in their reports.

Please provide your written response to the clarification questions by **5pm Wednesday 13 April 2016**. Your response and any supporting documents should be uploaded to NICE Docs/Appraisals.

Two versions of your written response should be submitted; one with academic/commercial-in-confidence information clearly marked and one with this information removed.

Please underline all confidential information, and separately highlight information that is submitted as commercial in confidence in turquoise, and all information submitted as academic in confidence in yellow.

If you present data that are not already referenced in the main body of your submission and that are academic/commercial in confidence, please complete the attached checklist for confidential information.

Please do not embed documents (PDFs or spreadsheets) in your response because this may result in them being lost or unreadable.

If you have any queries on the technical issues raised in this letter, please contact Joanne Holden, Technical Adviser (joanne.holden@nice.org.uk). Any procedural questions should be addressed to Stephanie Yates, Project Manager (Stephanie.yates@nice.org.uk) .

Yours sincerely

Nicola Hay
Technical Adviser – Appraisals

NICE National Institute for Health and Care Excellence

Level 1A
City Tower
Manchester
M1 4BT
United Kingdom

+44 (0)300 323 0140

On behalf of:

Dr Frances Sutcliffe
Associate Director – Appraisals
Centre for Health Technology Evaluation

Encl. checklist for confidential information

Section A: Clarification on effectiveness data

Literature searching

A1. **Priority question**: Please confirm the definition of moderate to severe RA applied in the inclusion criteria in the search strategy (Section 4.1.3.2 Population, page 44 of the company submission).

A2. Section 8.3.1 of the company submission states that non-RCT evidence was included for certolizumab pegol only but also notes that an RCT filter was applied to the main clinical effectiveness review. Please clarify the searching and study selection processes that led to the inclusion of the three observational studies presented as safety evidence (Yun *et al..,* Simard *et al.,* and Curtis *et al.*).

A3. It is noted that the company did not conduct an original search for health related quality of life evidence, instead using data from a review by Pennington & Davis (based on a simple Medline search conducted in 2012). Please provide the rationale for deciding not to update the review by Pennington & Davis and discuss the implications of not doing so.

A4. Please acknowledge the sources of the filters used to identify RCTs and economic studies, providing citations to peer-reviewed publications demonstrating their sensitivity/specificity where available.

A5. Table 8 (page 44 of the company submission) states that non-English language studies were excluded from the search strategy. Please confirm whether any potentially relevant non-English language RCTs for either certolizumab pegol, or any of the comparators were excluded from the submission.

Study Selection

A6. **Priority question:**

NICE National Institute for Health and Care Excellence

Level 1A
City Tower
Manchester
M1 4BT
United Kingdom

+44 (0)300 323 0140

a. Please clarify why RCT RA0025 (Kang *et al.,* 2012. Efficacy and safety of certolizumab pegol with concomitant methotrexate in Korean rheumatoid arthritis (RA) patients with an inadequate response to MTX. Annals of the Rheumatic Diseases; 71:666) was not included in the submission.

b. Please provide all available data relating to the biologic-experienced subgroup of patients from the above trial.

A7. **Priority question:**

a. Please clarify why the following comparator RCTs were excluded from the company submission: i) LITHE (tocilizumab), ii) ORAL Standard (adalimumab).

b. If the LITHE (tocilizumab) and ORAL Standard (adalimumab) RCTs are considered appropriate to include in the network meta-analysis, please update the analysis and provide revised estimates of effectiveness.

A8. Page 50 of the company submission states that the DOSEFLEX trial was followed by the DOSEFLEX II study (NCT00753454). However the data from DOSEFLEX II were not presented. Please provide all available data from DOSEFLEX II relevant to the biologic-experienced population.

Baseline characteristics

A9. **Priority question:** For each of the included certolizumab pegol RCTs, please provide all available details for:

a. receipt of prior biologic treatments (including type and duration of treatments),

b. reasons for discontinuation of prior biologic treatments.

A10. For each of the included certolizumab pegol RCTs, please provide the number (n/N, %) of patients who were from centres in the UK.

A11. Please clarify why the baseline characteristics data for the patients in the placebo group in the DOSEFLEX RCT could not be reported [Table 13 (page 73)]. Please provide the data if available.

Data analyses

A12. **Priority question:** Page 111 of the company submission states that the SWITCH trial data (data for Week 24) included in Figure 39 were re-drawn from the manuscript and that data point values were not available. Please confirm that these data are not available and if available, provide any supporting data.

NICE National Institute for Health and Care Excellence

Level 1A
City Tower
Manchester
M1 4BT
United Kingdom

+44 (0)300 323 0140

A13. Please describe the data imputation methods applied in the SWITCH trial (Section 4.4.1 Data imputation, page 62).

A14. Please provide any additional data from the included certolizumab pegol trials for the biologic-experienced population relating to the outcomes of HAQ-DI, health-related quality of life, pain and fatigue.

A15. Please state whether the occurrence of any adverse events in the included certolizumab pegol RCTs was statistically significant for patients receiving certolizumab pegol compared with placebo in i) the overall population, and ii) the anti-TNF experienced population.

A16. Safety data were not included in the company submission for the J-RAPID and HIKARI trials (Section 4.12.1 Adverse reactions in REALISTIC, DOSEFLEX, PREDICT, SWITCH).

a. Please clarify why safety data were not included for the J-RAPID and HIKARI trials,

b. Please provide safety data for the J-RAPID and HIKARI trials in i) the overall population, and ii) the anti-TNF experienced population.

Study design

A17. **Priority question:** Please clarify whether any of the included RCTs for certolizumab pegol included early escape or dose adjustments. If any of the RCTs included early escape or dose adjustment, please state how data were imputed to account for this.

A18. Please clarify why a run-in period was necessary for the DOSEFLEX study but not for other trials. In addition, please clarify the length of the run-in and what changes to patients' treatment took place.

A19. Please clarify why the sample size of the SWITCH trial was adjusted for potential drop-outs while the sample size of other trials was not adjusted for drop-outs.

Meta-analyses

A20. **Priority Question**: Throughout: Please provide estimates and 95% credible intervals for between-study standard deviations where these have not been provided.

A21. **Priority Question**: Please clarify why approximations of results using marginal univariate normal distribution or multivariate normal distributions were preferred to using draws from the joint posterior distribution (i.e. CODA) which would have maintained correlation.

NICE National Institute for Health and Care Excellence

Level 1A
City Tower
Manchester
M1 4BT
United Kingdom

+44 (0)300 323 0140

A22. **Priority Question**: Section 5.3.1.1: Please clarify how the "No response" rate for the reference treatment has been determined. As specified, "mu" is represents a study-specific parameter not a parameter for the general population.

A23. **Priority Question**: Table 67 (page 180 of the company submission): Please clarify where the means and standard deviations for the "No response" and cut-off came from.

A24. Section 4.9.2.1 (page 119 of the company submission) - Heterogeneity: Please clarify what attempt was made to explain the heterogeneity between studies and what the predictive distribution is for the effect in a new study.

A25. Section 4.9.3 (page 120 of the company submission) - Heterogeneity: Please clarify what attempt was made to explain the heterogeneity between studies and what is the predictive distribution is for the effect in a new study.

A26. Section 4.10.3 (page 123 of the company submission): Please clarify in what sense the results of Combe 2012 are implausible and provide an explanation for why this might be the case. Please clarify why the results from Combe 2012 were deemed implausible (p123).

A27. Please clarify why Combe 2012 was not included in Figure 62 (page 189 of the company submission).

A28. Section 4.10.5 (page 124 of the company submission) states ‘An adjusted indirect analysis method was chosen over Bayesian for outcomes where evidence network included not more than two competing interventions. For evidence networks assessing more than two competing interventions a Bayesian NMA was performed’.

a. Please clarify why the distinction was made in the method of analysis and discuss what implication this has for issues such as dealing with heterogeneity and subsequent inference.

b. Please clarify why Bucher analyses were preferred to a Bayesian Random effects model: if possible re-analyse using a Bayesian Random Effects Model accounting for heterogeneity.

A29. Section 4.10.5.2 (pages 126-127 of the company submission): Please clarify why there are no ACR70 data according to the following sentence whereas there are ACR20 and ACR50 data (NOTE: zeroes are classed as data): Indirect analysis was not possible for ACR70 response at 3-months due to there being no data from the comparator studies.

NICE National Institute for Health and Care Excellence

Level 1A
City Tower
Manchester
M1 4BT
United Kingdom

+44 (0)300 323 0140

A30. Section 4.10.6.1 (page 128-129 of the company submission): Please confirm that the results are random effects mean and not predictive distributions for the effects in new studies.

A31. Section 4.10.6.1 (page 128 of the company submission): Please clarify what is considered to be a clinically relevant non-inferiority margin for claims of “at least as effective”.

A32. Section 5.3.1.1 (page 177 of the company submission): Please clarify what is meant by, "These parameters were varied using a series of univariate normal distributions."

A33. Section 5.3.1.5 (pages 189-193): Please clarify the following:

a. The expected magnitude of the between study variance given the perceived heterogeneity.

b. Why a Random Effects model incorporating this was not used.

A34. Table 69 (page 192 of the company submission): Please clarify whether the ‘Trial-specific baseline parameter’ refers to the baseline response rate for the reference treatment.

A35. Section 5.3.2 (pages 193-195): Please clarify whether interaction effects were assessed.

A36. Throughout the clinical effectiveness submission: Please provide p-values where p-values have not been calculated (for example Table 16, page 83 of the company submission) these have not been provided. Where possible, please also provide data and statistical significance for combination and monotherapy subgroups.

Section B: Clarification on cost-effectiveness data

B1. **Priority Question:** It is noted that other bDMARDs are only recommended for severe patients (defined as a DAS28 score > 5.1) after the failure of a TNF inhibitor (TNFi). It is further noted that the company submission does not include an assessment of the clinical and cost effectiveness of certolizumab pegol in moderate to severe RA. Please clarify whether the ICERs presented should therefore be assumed to be applicable to severe patients only, particularly given that the patient characteristics in the model are based on the TNFi experienced subgroup of REALISTIC, which has a mean DAS28 score of [redacted] If this is not the case please provide an ICER for the moderate to severe RA population against the appropriate comparator.

B2. **Priority Question:** Please clarify why certolizumab pegol was considered for population A as an addition to the treatment sequence before rituximab instead of as

Level 1A
City Tower
Manchester
M1 4BT
United Kingdom

+44 (0)300 323 0140

an alternative to rituximab. It is noted that this elongates the sequence. Please clarify why, if elongated sequences are being considered certolizumab pegol was not positioned after rituximab and an incremental analysis performed between these two strategies.

B3. **Priority Question:** Please clarify why abatacept was included after tocilizumab in the treatment sequence for population A (Table 64, page 174 of the company submission). This appears to be outside of NICE's recommendations for abatacept.

B4. **Priority Question**: Please provide a scenario analysis where mortality is not affected by changes in the HAQ score and is only dependent on baseline HAQ, as per the Assessment Group's assumptions in NICE technology appraisal 375 'Adalimumab, etanercept, infliximab, certolizumab pegol, golimumab, tocilizumab and abatacept for rheumatoid arthritis not previously treated with DMARDs or after conventional DMARDs only have failed'.

B5. **Priority Question:** The values for changes in HAQ score from RADIATE (-0.39 for biologics, -0.05 for methotrexate) are presumed to be values for the entire population. It is anticipated that the HAQ improvements for good and moderate responders would be greater than for the entire population. Please clarify whether this has been considered, noting that 32.3% and 83.5% of patients were non-responders in the tocilizumab and methotrexate groups respectively. In addition, please clarify why no analyses were run assuming the same HAQ change based on EULAR response independent of treatment.

B6. Please clarify why certolizumab pegol was not considered for population A instead of tocilizumab in patients with an inadequate response to rituximab.

B7. Please clarify why non-biologics were included as a combined therapy for population A and as individual lines of therapy for population B and C.

B8. Please clarify why a simple linear regression was used to model the relationship between HAQ-DI and pain (Figure 64, page 195 of the company submission) instead of a quadratic model as in NICE technology appraisal 375.

B9. Please clarify why for the discontinuation of second and subsequent treatments a exponential distribution was used instead of the Weibull distribution used for the first treatment. Please provide the justification as to why the transition probability for second and subsequent treatments were assumed to equal that for the Weibull distribution between months 6 and 12.

B10. Please clarify why the mapping between HAQ and EQ-5D from Brennan et al. was used in the base case analysis instead of the more recent one used within the

NICE National Institute for Health and Care Excellence

Level 1A
City Tower
Manchester
M1 4BT
United Kingdom

+44 (0)300 323 0140

Assessment Group model for NICE technology appraisal 375 (i.e. Hernandez-Alava et al.).

B11. Please clarify why the implementation of Hernandez-Alava et al. utility mappings only includes a subset of the covariates, and why the distributions for each of the classes of the mixture model are independently applied instead of combined through a weighted average (using the class membership probabilities as weights). In addition, please clarify to what extent inaccuracy within the calculation of utility values would impact on the resultant ICERs.

B12. Please clarify why a Dirichlet distribution was used for the weight parameter for the PSA rather than fitting a distribution. This may have an impact should the number of vials required change within a weight band.

B13. Please clarify why the cycle length is different (3 months instead of 6 months) for the second and third cycles of the model. Please confirm whether there is an error in cycles 2 and 3 with the 6-month discontinuation probabilities being applied to 3 monthly time cycles.

B14. Please clarify why nurse visits for a percentage of patients unwilling or unable to perform subcutaneous injections were not considered as in the Assessment Group's economic model for NICE technology appraisal 375.

B15. Please clarify the following:

a. Why the half-cycle correction implementation counts the utility of the first cycle 1.5 times instead of 0.5 times as it is customary when applying the half-cycle correction (or just once, if the first cycle should be exempt of half-cycle correction).

b. Why in some cases (e.g. Markov_C!CK28) the half cycle correction mixes utility values from two different states.

c. Why in some cases (e.g. Markov_C!CK29) the utility value of a single cycle are halved instead of calculating the average of two subsequent cycles.

B16. Please clarify why the 80 mg dose of tocilizumab was not considered in the model.

B17. Please clarify why in Figure 57 (page 172 of the company submission) the arrows from the two "Continue Tx HAQ change = d" states in the fourth row ("Cycle 3+;time=1.5 years") point at the "FU treatment 1* (**2nd** cycle+)" state instead of "FU treatment 1* (**1st** cycle)".

B18. Please clarify the apparent discordance between the percentage of good or moderate responders reported by Emery et al. in the methotrexate treatment arm

NICE National Institute for Health and Care Excellence

Level 1A
City Tower
Manchester
M1 4BT
United Kingdom

+44 (0)300 323 0140

(16.5%) and the probability of inadequate response (83.7%) reported for the RADIATE trial in the company submission (page 200). It is believed the addition of the two values should equal 100%.

B19. Please clarify why utility values were not age-adjusted.

B20. Please clarify why biosimilars for etanercept were not considered.

B21. Please provide a sensitivity analysis using the change in HAQ associated with good, moderate and no EULAR response as used in NICE technology appraisal 375. These values are 0.672, 0.317 and 0 respectively.

B22. Please provide a sensitivity analysis in relation to the costs of palliative care.

Section C: Textual clarifications and additional points

C1. Please clarify whether the phrase waste within “unopened vials” being lost is a typographical error. From the model it would appear that leftover drugs in open vials are assumed to be lost, rather than unopened vials. Please clarify why the budget impact model assumes no vial wastage in contrast to the economic model.

C2. Table 13 (page 73 of the company submission): Please confirm that the percentages within the TNFi experienced patients sub-row in the REALISTIC study are incorrect as they have been divided by the wrong number of patients. There are several other instances within Table 13 of apparent miscalculation, please check the values throughout this table.

C3. Table 24 (page 109 of the company submission): The percentage of patients in remission in the certolizumab pegol treatment arm appears to be incorrectly calculated, please confirm if this is correct.

C4. Please update Figure 2 (page 41 of the company submission) so that NICE technology appraisal 375 is taken into consideration. This would include renaming some of the technology appraisals and adding tocilizumab monotherapy as a first-line bDMARD option. This latter point may mean that the heading of one box may need to be changed.

C5. Rendas-Baum et al. report that more than 90% of rheumatologists switched patients to alternative TNFi therapy yet page 38 suggests 94%. Please clarify from where the more precise value was identified.

C6. Please clarify whether the mean length of the first anti-TNFα therapy in Hyrich et al. was 13 months rather than the 6 months reported in page 38 of the company submission.

NICE National Institute for Health and Care Excellence

Level 1A
City Tower
Manchester
M1 4BT
United Kingdom

+44 (0)300 323 0140

C7. Please clarify whether 16% of patients who discontinued first biologic agent due to inefficacy also discontinued second biologic agent due to inefficacy, as reported in Hyrich et al, instead of 13% as reported on page 38 of the company submission.

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

Single Technology Appraisal

Certolizumab Pegol for Moderate to Severe Rheumatoid Arthritis Inadequately Responding to a Prior TNFi [ID824]

17th April 2016

UCB Response to ERG Clarification Questions

Underlined confidential information, and separately highlighted information that is submitted as 'commercial in confidence' in turquoise and all information submitted as 'academic in confidence' in yellow.

UCB Response to ERG Clarification Questions

UCB welcomes the opportunity to respond to the questions from the Evidence Review Group (ERG) and the NICE Technical Team, following their initial review the single technology appraisal (STA) submission for certolizumab pegol (CZP) for the treatment of moderate to severe rheumatoid arthritis (RA) that as inadequately responded to a prior TNF inhibitor (TNFi) [ID824]. UCB encloses its responses and further clarification to these questions below.

A. Clarification on effectiveness data

Literature searching

A.1 **Priority question:** Please confirm the definition of moderate to severe RA applied in the inclusion criteria in the search strategy (Section 4.1.3.2 Population, page 44 of the company submission).

UCB response:

Moderate to severe disease activity is defined as disease activity score 28 (DAS28)>3.2. Please also refer to the study population described in Table 1 (Section 1.1, page 18) of UCB's submission.

A.2 Section 8.3.1 of the company submission states that non-RCT evidence was included for certolizumab pegol only but also notes that an RCT filter was applied to the main clinical effectiveness review. Please clarify the searching and study selection processes that led to the inclusion of the three observational studies presented as safety evidence (Yun et al.., Simard et al., and Curtis et al.).

UCB response:

The study design filter applied in the searches was a very broad filter to capture both randomised controlled trial (RCT) and non-RCT studies with key terms searched as MeSH terms such as 'clinical study'; 'clinical trial'; 'controlled study'; 'prospective study'; and 'comparative study' to identify non-RCTs. As the objective of the clinical review and network meta-analysis (NMA) was to assess the efficacy of CZP versus other TNFis, these three observational studies were not discussed in the review section. Of these three studies, one study was identified through searches (Simard et al),[1] one study was identified through hand searching but not included due to study design and only safety information (Curtis et al).[2] The final study, (Yun et al.) was not captured in initial searches since it was indexed after the search date (November 2015) and was identified through additional horizon scanning conducted closer to the submission deadline to capture additional safety data distinct from the clinical and economic-focussed systematic searches.[3]

A.3 It is noted that the company did not conduct an original search for health related quality of life evidence, instead using data from a review by Pennington & Davis (based on a simple Medline search conducted in 2012). Please provide the rationale for deciding not to update the review by Pennington & Davis and discuss the implications of not doing so.

UCB response:

The review by Pennington and Davis was validated against the EQ-5D mapping database published by the Health Economics Research Centre (HERC) at the Nuffield Department of Public Health (available from: http://www.herc.ox.ac.uk/downloads/herc-database-of-mapping-studies).

According to the HERC website (accessed on 8th April 2016), the latest update of the mapping database was on 30th September 2015. The database is expected to contain all contemporary studies that report the mapping of HAQ to EQ-5D.

The HERC database lists 21 publications that report algorithms mapping HAQ to EQ-5D. All studies were published between the years of 1997 and 2012[1]. None of the algorithms were published in the period between the search by Pennnington and Davis (2012), and the most recent HERC search (September 2015). As such, an update to the review by Pennington and Davis would not yield further evidence on the relationship between HAQ and EQ-5D in patients with RA, and hence was not considered necessary for the purpose of this submission.

Of the 21 publications listed in the HERC database, 14 were included in the review by Pennington and Davis. Of the 7 excluded studies, 1 study was linked to a second study that was included in Pennington and Davis, and a further study reported the mapping of HAQ to EQ-5D in a heterogeneous population that included patients without RA. The remaining 5 studies were published prior to 2012. A summary of all 21 studies identified in the HERC database are provided in Appendix 1 to this response document.

A.4 Please acknowledge the sources of the filters used to identify RCTs and economic studies, providing citations to peer-reviewed publications demonstrating their sensitivity/specificity where available.

UCB response:

Standardised filters were used for identifying the RCTs and economic studies. These filters were adapted from the SIGN filters. Appropriate SIGN filters can be accessed at the following website:

- http://www.sign.ac.uk/methodology/filters.html

A.5 Table 8 (page 44 of the company submission) states that non-English language studies were excluded from the search strategy. Please confirm whether any potentially relevant non-English language RCTs for either certolizumab pegol, or any of the comparators were excluded from the submission.

UCB response:

No relevant non-English language RCT for CZP or any other comparator were identified.

Study Selection

A.6 Priority question:

a. Please clarify why RCT RA0025 (Kang et al., 2012. Efficacy and safety of certolizumab pegol with concomitant methotrexate in Korean rheumatoid arthritis (RA) patients with an inadequate response to MTX. Annals of the Rheumatic Diseases; 71:666) was not included in the submission.

UCB response:

The RCT RA0025 (Kang et al., 2012. Efficacy and safety of certolizumab pegol with concomitant methotrexate in Korean rheumatoid arthritis (RA) patients with an inadequate response to MTX. Annals of the Rheumatic Diseases; 71:666) was not included in the original UCB submission. Out of the 121 patients included in the Full Analysis Set (FAS), only 18 patients (14.9%; 7 [17.5%] in the placebo (PBO) with methotrexate (MTX) group and 11 [13.6%] treated with CZP 200 mg Q2W and MTX) were previously exposed to a prior TNFi. The pre-specified subgroup analysis of the primary endpoint (ACR20 at Week 24 is provided below).

b. Please provide all available data relating to the biologic-experienced subgroup of patients from the above trial.

UCB response:

[1] One article was published in 2013 – however, this study was linked to an earlier publication dated 2012 and included in Pennington and Davis.

Table A: RA0025 study: ACR20 response rates at Week 24 (TNFi experienced population, non-responder imputation (NRI)

	PBO + MTX [redacted]	CZP 200 mg Q2W + MTX [redacted]
ACR20 Response, n (%)		
Week 24	[redacted]	[redacted]

FAS; NRI. No OR or p-values calculated given the small sample size, as per study protocol.

A.7 **Priority question**:

a. Please clarify why the following comparator RCTs were excluded from the company submission: i) LITHE (tocilizumab), ii) ORAL Standard (adalimumab).

UCB response:

Both LITHE and ORAL standard studies were identified in the literature searches and were excluded during the screening for following reason:

- The LITHE study (tocilizumab) was conducted in patients with inadequate response to MTX therapy and this study did not provide biological disease-modifying anti-rheumatic drug (bDMARD)-experienced data.

- Similarly, the ORAL standard study (adalimumab) excluded patients who failed any prior TNFi for either lack of efficacy or a TNFi mechanism-related adverse event (AE). Although 7.8% of the patients randomised to receive adalimumab (16 of 204 patients) had previously received a TNFi, however subgroup data for these patients was not reported in the study publication.

Thus, these two studies were not considered for inclusion in the review.

b. If the LITHE (tocilizumab) and ORAL Standard (adalimumab) RCTs are considered appropriate to include in the network meta-analysis, please update the analysis and provide revised estimates of effectiveness.

UCB response:

Not applicable. Please refer to explanation in A.7a.

A.8 Page 50 of the company submission states that the DOSEFLEX trial was followed by the DOSEFLEX II study (NCT00753454). However the data from DOSEFLEX II were not presented. Please provide all available data from DOSEFLEX II relevant to the biologic-experienced population.

UCB response:

DOSEFLEX II (NCT00753454) was a Phase 3b, multicenter, open-label, follow-up study to DOSEFLEX (NCT00580840) designed to continue to assess the safety and efficacy of CZP. Subjects having completed the Week 34 assessment in DOSEFLEX or having met the predefined criteria for flare (defined as subjects that were randomized at Week 18 and experienced an equal to Baseline [Week 0] or worsened SJC and TJC at 2 consecutive visits between Week 18 and Week 34 inclusive), were given the option to enroll in DOSEFLEX II and receive CZP 400mg at entry, Week 2, and Week 4 followed by CZP 200mg Q2W in combination with MTX until the drug was commercially available for the indication of RA in the subject's country or region or until further notice from UCB.

Data from the pooling of the DOSEFLEX study and its open-label extension DOSEFLEX II are presented in Table B, Table C and Table D, and report ACR and EULAR responses and HAQ-DI scores, respectively, in patients randomised in DOSEFLEX to either the CZP 200 mg Q2W or CZP 400 mg Q4W groups (combined) from Week 18 onwards. Data are reported throughout the 34 weeks of DOSEFLEX, and into the open label extension (DOSEFLEX II) up to Week 42; data are not presented beyond this time point, due to the substantial drop in patient numbers due to study site closure after approval of CZP in the US. The results showed that the initial clinical response and improvements in HAQ-DI with CZP were maintained throughout Week 42 in the open-label.

Table B: Pooling of DOSEFLEX and DOSEFLEX II studies: ACR response rates to Week 42 (TNFi-experienced population, randomised to CZP from Week 18 onwards in DOSEFLEX, Open label set, NRI)

	CZP + MTX* [redacted]
ACR20, n (%)	
Week 4	
Week 16	
Week 34	
Week 42	
ACR50, n (%)	
Week 4	
Week 16	
Week 34	
Week 42	
ACR70, n (%)	
Week 4	
Week 16	
Week 34	
Week 42	

Open label set (NRI); * CZP+MTX includes all patients randomised in DOSEFLEX to either the CZP 200 mg Q2W or CZP 400 mg Q4W groups (combined) from Week 18 onwards.

Table C: Pooling of DOSEFLEX and DOSEFLEX II studies: EULAR response rates to Week 42 (TNFi-experienced population, randomised to CZP from Week 18 onwards in DOSEFLEX, Open label set, LOCF)

		CZP + MTX*
EULAR Response, n (%)		
Week 4 [redacted]	Good	
	Moderate	
	None	
Week 16 [redacted]	Good	
	Moderate	
	None	
Week 34 [redacted]	Good	
	Moderate	
	None	
Week 42 [redacted]	Good	
	Moderate	
	None	

Open label set (LOCF); * CZP+MTX includes all patients randomised in DOSEFLEX to either the CZP 200 mg Q2W or CZP 400 mg Q4W groups (combined) from Week 18 onwards.

Table D: Pooling of DOSEFLEX and DOSEFLEX II studies: Mean changes in HAQ-DI scores to Week 42 (TNFi-experienced population, randomised to CZP from Week 18 onwards in DOSEFLEX, Open label set, LOCF)

	CZP + MTX* [redacted]
HAQ, mean score (SD) [mean change from baseline]	
Week 0 (Baseline)	
Week 4	
Week 16	
Week 34	
Week 42**	

Open label set (LOCF); * CZP+MTX includes all patients randomised in DOSEFLEX to either the CZP 200 mg Q2W or CZP 400 mg Q4W groups (combined) from Week 18 onwards. **n=64 at Week 42.

Baseline characteristics

A.9 **Priority question:** For each of the included certolizumab pegol RCTs, please provide all available details for:

a. Receipt of prior biologic treatments (including type and duration of treatments),

UCB response:

A summary of the baseline characteristics in terms of prior TNFis use in the included CZP studies, REALISTIC, DOSEFLEX, PREDICT, J-RAPID and HIKARI can be found below. This presents the proportion of most commonly used prior TNFis (Table E), in addition to the mean duration of prior TNFi treatment (Table F). Of note, the duration of prior treatment for REALISTIC and DOSEFLEX could not be estimated due [redacted]

Table E: Baseline characteristics: summary of prior TNFi exposure in CZP studies‡

	REALISTIC†		DOSEFLEX‡			PREDICT†	J-RAPID†		HIKARI†	
	CZP 200 mg Q2W (n=851)	PBO (n=212)	CZP 200 mg Q2W (n=70)	CZP 400 mg Q4W (n=70)	PBO (n=69)	CZP 200 mg Q2W (n=733)	CZP 200 mg Q2W (n=82)	PBO (n=77)	CZP 200 mg Q2W (n=116)	PBO (n=114)
Proportion of patients with prior TNFi exposure, n (%)										
Adalimumab	[redacted]									
Etanercept	[redacted]									
Golimumab	[redacted]									
Infliximab	[redacted]									

†Full analysis set; ‡Enrolled set

‡Data presented in this table refers to: past TNFis (REALISTIC), prior TNFis (DOSEFLEX, PREDICT) and pre-treatment (J-RAPID and HIKARI). Past TNFi refers as any TNFi taken in the past in REALISTIC. Prior TNFi refers to any TNFi taken in the past in DOSEFLEX and PREDICT. Pre-treatment refers to any anti-rheumatic treatments taken in the past in J-RAPID and HIKARI.

Table F: Baseline characteristics: summary of duration of exposure for prior TNFis in CZP studies‡

	PREDICT†	J-RAPID†		HIKARI†	
	CZP 200 mg Q2W (n=733)	CZP 200 mg Q2W (n=82)	PBO (n=77)	CZP 200 mg Q2W (n=116)	PBO (n=114)
Duration of exposure, mean days (SD)					
Adalimumab	[redacted]				
Etanercept	[redacted]				
Golimumab	[redacted]				
Infliximab	[redacted]				

†Full analysis set; ‡Enrolled set

‡Data presented in this table refers to: past TNFis (REALISTIC), prior TNFis (DOSEFLEX/PREDICT) and pre-treatment (J-RAPID and HIKARI). Past TNFi refers as any TNFi taken in the past in REALISTIC. Prior TNFi refers to any TNFi taken in the past in DOSEFLEX and PREDICT. Pre-treatment refers to any anti-rheumatic treatments taken in the past in J-RAPID and HIKARI.

b. Reasons for discontinuation of prior biologic treatments.

UCB response:

Table G reports the baseline characteristics with respect to reasons for discontinuation of prior TNFis and other biologics in REALISTIC, DOSEFLEX and PREDICT._It was not possible to generate these data for J-RAPID and HIKARI, since they were not collected.

Table G: Baseline characteristics: reasons for discontinuation of prior TNFis and other biologics in REALISTIC, DOSEFLEX and PREDICT studies‡

	REALISTIC†		DOSEFLEX‡			PREDICT†
	CZP 200 mg Q2W (n=851)	PBO (n=212)	CZP 200 mg Q2W (n=70)	CZP 400 mg Q4W (n=70)	PBO (n=69)	CZP 200 mg Q2W (n=733)
Overall, n (%)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Primary lack of response, n (%)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Secondary loss of response, n (%)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Loss of efficacy	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Intolerance/Adverse drug reaction, n (%)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Partial response, n (%)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Financial, n (%)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Study ended, n (%)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Other, n (%)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Unknown, n (%)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

†Full analysis set; ‡Enrolled set; N/A: not applicable (categories not assessed in the study).
‡The definition of category of reasons of discontinuation for past treatment was not the same in the REALISTIC, DOSELEX and PREDICT therefore data from the trials cannot be directly compared. Category names were identical in REALISTIC and PREDICT however definition of each category may differ between the two trials. Category names were different in DOSELEX compared to REALISTIC/PREDICT.

A.10 For each of the included certolizumab pegol RCTs, please provide the number (n/N, %) of patients who were from centres in the UK.

UCB response:

None of the submitted CZP RCTs (REALISTIC, DOSEFLEX, PREDICT, SWITCH, J-RAPID and HIKARI), included participants recruited from centres in the UK.

As indicated in the UCB submission (Table 11, page 53) the above mentioned studies were conducted in the following countries (full details available on ClinicalTrials.gov):

- REALISTIC: USA, Canada and Europe (France, Germany, Italy, The Netherlands, and Spain)
- DOSEFLEX: USA, Canada and France
- PREDICT and SWITCH: US only
- J-RAPID and HIKARI: Japan only

A.11 Please clarify why the baseline characteristics data for the patients in the placebo group in the DOSEFLEX RCT could not be reported [Table 13 (page 73)]. Please provide the data if available.

UCB response:

Baseline characteristic data for the TNFi-experienced PBO population in DOSEFLEX were not available at the time of submission. Please find the missing data in Table H below. For completeness we have also included the information included in the UCB submission for this group (ie DAS28(ESR) and RF-positive patients, as per Table 13, page 73 of the submission document).

Table H: DOSEFLEX study: Baseline characteristics of previously TNFi-exposed (PBO+MTX arm)

	DOSEFLEX TNFi-experienced PBO + MTX† (n=29)
TNFi experienced of group, n (%)‡	[redacted]
Mean age (SD), years	[redacted]
Female, n (%)	[redacted]
Mean disease duration, years (SD)	[redacted]
HAQ-DI mean (SD)	[redacted]
DAS28(ESR), mean (SD)‡	[redacted]
RF-positive (≥14 IU/mL), n (%)‡	[redacted]

†For DOSEFLEX, the PBO stratification data represent patients who completed the run-in phase and were subsequently randomised into the three treatment groups (PBO, CZP 200 mg Q2W or CZP 400 mg Q4W) for the double-blind phase.
‡Values provided in Table 13, page 73 of UCB's submission document

Data analyses

A.12 **Priority question**: Page 111 of the company submission states that the SWITCH trial data (data for Week 24) included in Figure 39 were re-drawn from the manuscript and that data point values were not available. Please confirm that these data are not available and if available, provide any supporting data.

UCB response:

The SWITCH study was an investigator-initiated study (IIS), and UCB does not have access to the trial data. The graph was re-drawn from the manuscript for illustrative purposes; the individual data point for the Week 12 ACR20 response rate was written in the text of the manuscript,[4] and therefore presented as a data point on the graph in UCB's submission (Figure 39, page 111). However, ACR50/70 at Week 12 and all response rates at Week 24 were not published, and were instead estimated using graph-reading software to complete the graph. UCB does not have exact numbers to provide; indeed the ACR50/70 response rates at Week 12 were reported in error and the exact values should be disregarded.

A.13 Please describe the data imputation methods applied in the SWITCH trial (Section 4.4.1 Data imputation, page 62).

UCB response:

As per UCB's response to question A.12 above, given the SWITCH is an IIS, UCB does not have access to information beyond what is provided in the manuscript, which is that a Cochran-Mantel-Haenszel $\chi 2$ test was used for the CZP vs PBO comparison. UCB cannot provide further information on the data imputation methods.[4]

A.14 Please provide any additional data from the included certolizumab pegol trials for the biologic-experienced population relating to the outcomes of HAQ-DI, health-related quality of life, pain and fatigue.

UCB response:

The data from the CZP studies that were provided are summarised in Section 4.2.1 (Table 10, page 49, of UCB's submission). These include HAQ-DI, pain, fatigue and HRQoL (kindly note that pain was reported as a component of the ACR response criteria, and can be found in UCB's submission

[Appendices 8.7.3 and 8.8.3]). As indicated in UCB's submission, there were some differences in the measures evaluated across the studies and not all endpoints were available in all studies.

A.15 Please state whether the occurrence of any adverse events in the included certolizumab pegol RCTs was statistically significant for patients receiving certolizumab pegol compared with placebo in i) the overall population, and ii) the anti-TNF experienced population.

UCB response:

The included RCTs were not powered to detect safety differences, therefore statistical comparisons of safety events between patients receiving CZP and those receiving PBO were not generated for either the overall population or the TNFi experienced population. Moreover, the majority of safety events tend to be infrequent, thus making it impossible to power studies for these unknown signals. For this reason UCB with CZP are part of registries including the British Society for Rheumatology Biologics Register (BSRBR) in order to ensure comparable data with other TNFi therapies. Comparability of safety data can be conducted using data from these registries and additionally, real-world data from the Centres for Medicare and Medicaid Services Chronic Condition Data Warehouse was analysed by Yun et al. 2015 and presented in the submission (Section 4.12.3.1, page 155).[3] This study reported no increase in risk of hospitalised infection in patients switched to CZP compared to patients switched to abatacept. Additionally, data from the Swedish Biologics Register (ARTIS) was presented in the submission (Section 4.12.3.2, page 157).[1] Simard et al. 2011 showed that history of infection at the start of a second biological disease-modifying anti-rheumatic drug (bDMARD) was greater among those who switched to a non-TNFi regimen compared with those initiating a second TNFi therapy (32 vs 18%).[1] A large pooled analysis of safety data from 10 completed RCTs and several open-label extensions on CZP in RA was also presented in the submission (Section 4.12.2, page 155) and this analysis did not identify any new or unexpected safety signals associated with CZP.[5]

A.16 Safety data were not included in the company submission for the J-RAPID and HIKARI trials (Section 4.12.1 Adverse reactions in REALISTIC, DOSEFLEX, PREDICT, SWITCH).

a. Please clarify why safety data were not included for the J-RAPID and HIKARI trials,

UCB response:

The safety data for J-RAPID and HIKARI are provided below. It is important to note that Japanese patients report AEs differently to non-Japanese patients due to cultural and lifestyle differences. Additionally, some opportunistic infections (OIs) are defined differently compared to the rest of the world. Caution should therefore be made when comparing safety data from studies conducted in Japanese patients versus non-Japanese patients.

b. Please provide safety data for the J-RAPID and HIKARI trials in i) the overall population, and ii) the anti-TNF experienced population.

UCB response:

J-RAPID

The overall incidence of AEs amongst the entire trial population was comparable between the CZP 200 mg Q2W in combination with MTX and PBO in combination with MTX groups in the J-RAPID study (76.8% vs 66.2%, respectively; Table I). The majority of AEs in both groups were of mild to moderate intensity. SAEs were infrequent, occurred in 4 (4.9%) of CZP 200 mg Q2W-treated patients and 1 (1.3%) patient in the PBO group. There were no cases of tuberculosis or malignant disease and no deaths during the 24 weeks of the study.[6]

The most frequent AE by system organ class was infections and infestations, occurring in 84 patients (35.1%) in the CZP-treated groups (including the CZP 200 mg Q2W group, and patients treated with CZP 100 mg Q2W and CZP 400 mg Q2W [note that CZP 100 mg Q2W and CZP 400 mg Q2W are unlicensed doses within the EU]) and 19 patients (24.7%) in the PBO group. Administration site reaction (2.5% and 0.0%), injection site erythema (0.8% and 0.0%), injection site hematoma (0.4% and 0.0%), injection site haemorrhage (0.0% and 1.3%), injection site mass (0.4% and 0.0%) and injection site reaction (0.4% and 1.3%) were reported across all CZP-treated groups and the PBO group,

respectively, and the majority of reactions were mild. In the CZP groups, the proportions of patients who showed high aspartate aminotransferase and alanine aminotransferase levels after administration despite normal values at baseline were slightly higher than in the PBO group.[6]

No safety analyses have been conducted specifically in the TNFi experienced population of the J-RAPID trial due to the very low cohort size (see response to A15).

Table I: Safety up to the end of the 24 week double-blind J-RAPID study (overall safety population)

Exposure and AEs	CZP 200 mg Q2W + MTX (n=82)	PBO + MTX (n=77)
Duration of exposure, PY	36.80	25.16
Any AEs by maximum intensity, n (%)	63 (76.8)	51 (66.2)
Mild	41 (50.0)	28 (36.4)
Moderate	20 (24.4)	22 (28.6)
Severe[a]	2 (2.4)	1 (1.3)
Treatment-related[b], n (%)	31 (37.8)	21 (27.3)
Serious AE (total), n (%)	4 (4.9)[c]	1 (1.3)
Malignancy, n (%)	0	0
Deaths, n (%)	0	0
Most common AEs[d] (≥5% in any group), n (%)		
Nasopharyngitis	11 (13.4)	9 (11.7)
Abnormal hepatic function	3 (3.7)	4 (5.2)
RA exacerbation	4 (4.9)	9 (11.7)
Pharyngitis	5 (6.1)	3 (3.9)
Serious AEs, n (%)		
RA	1 (1.2)	0
Bronchitis	1 (1.2)	0
Pyelonephritis	1 (1.2)	0
Purulent myositis	1 (1.2)	0
Subcutaneous tissue abscess	1 (1.2)	0
Urosepsis	1 (1.2)	0
Anal fistula	0	1 (1.3)

PY: patient years.
[a]Severe AE defined as an event that prevents work or daily activities.
[b]Treatment-emergent AEs for which the relationship to the study drug cannot be ruled out.
[c]6 events in 4 patients.
[d]Preferred terms according to MedDRA terminology.
Adapted from Yamamoto et al. (2014)[6]

HIKARI

Treatment-emergent AEs were reported in 71.6% (83/116) of patients treated with CZP 200 mg Q2W and 58.8% (67/114) of PBO patients, the majority being of mild to moderate intensity (Table J). AEs leading to withdrawal were more frequent in the CZP group. The most frequently reported AE by preferred term in both groups was nasopharyngitis. Skin rash was more frequent with CZP than PBO. Injection site erythema (3 patients, 2.6%), injection site reaction (3 patients, 2.6%), administration site reaction (2 patients, 1.7%), and injection site hematoma (1 patient, 0.9%) were reported in patients treated with CZP 200 mg Q2W. All of these reactions were mild. No administration site reactions were observed in the PBO group. Serious AEs were observed in 13 patients (14 events) in the CZP group and in three patients (5 events) in the PBO group. The most common Serious AE in both groups was infections (CZP 3.4% vs. PBO 0.9%). In the CZP group there were 4 events of serious infection including 1 event each of *pneumocystis jiroveci* pneumonia (PCP), pneumococcal pneumonia, herpes zoster and bacterial arthritis. In the PBO group there were 2 events of serious infection (1 event each of cellulitis and influenza), both occurring in the same patient. In the CZP group, 1 patient died of a rupture of a dissecting aortic aneurysm in the thoracic region, but this was considered unlikely to have been related to study medication. There were no cases of tuberculosis, but there was 1 report of malignant disease in the PBO group.[7]

No safety analyses have been conducted specifically in the TNFi experienced population of the HIKARI trial due to the very low cohort size (see response to A15).

Table J: Safety up to the end of the 24 week double-blind HIKARI study (overall safety population)

Exposure and AEs	CZP 200 mg Q2W (n=116)	PBO (n=114)
Duration of exposure, PY	49.43	34.08
Any AEs by maximum intensity, n (%)	83 (71.6)	67 (58.8)
Mild	33 (28.4)	29 (25.4)
Moderate	44 (37.9)	36 (31.6)
Severe[a]	6 (5.2)	2 (1.8)
Treatment-related, n (%)	44 (37.9)	24 (21.1)
Serious AE (total), n (%)	13 (11.2)[b]	3 (2.6)[c]
Deaths, n (%)	1 (0.9)	0
AEs leading to withdrawal	9 (7.8)	3 (2.6)
Most common AEs[d] (≥3% in any group), n (%)		
Nasopharyngitis	20 (17.2)	16 (14.0)
Rash	10 (8.6)	0
Pharyngitis	6 (5.2)	5 (4.4)
Eczema	6 (5.2)	3 (2.6)
RA	5 (4.3)	14 (12.3)
Abnormal hepatic function	4 (3.4)	4 (3.5)
Hypertension	4 (3.4)	1 (0.9)
Constipation	4 (3.4)	0
Upper respiratory tract infection	3 (2.6)	4 (3.5)
Serious AEs, n (%)	4 (3.4)	1 (0.9)[e]

PY: patient years.
[a]Severe AE defined as an event that prevents work or daily activities.
[b]14 events in 13 patients.
[c]5 events in three patients.
[d]Preferred terms according to MedDRA terminology.
[e]2 events in the same patient.
Adapted from Yamamoto et al. (2014)[7]

Study design

A.17 **Priority question**: Please clarify whether any of the included RCTs for certolizumab pegol included early escape or dose adjustments. If any of the RCTs included early escape or dose adjustment, please state how data were imputed to account for this.

UCB response:

As indicated in the UCB submission (Section 4.3, page 49), none of the included RCTs for CZP included dose adjustments of CZP. Only the 2 Japanese studies included an early escape criteria. A brief summary of the study designs is provided below (for full details please refer to UCB submission, Section 4.3):

- REALISTIC study: [redacted].
- PREDICT study: patients who failed treatment at Week 12 due to a lack of response (defined as no improvement in RAPID3 for those randomised to the RAPID3 assessment group, or ≤1 point improvement for those randomised to the CDAI assessment group) were withdrawn from PREDICT.[8]
- DOSEFLEX: ACR20 responders at Week 16 were randomised 1:1:1 to receive PBO, or one of either CZP maintenance doses, in combination with MTX, from Week 16. Patients not achieving an ACR20 response at Week 16 were withdrawn from the study from Week 16 and were not invited to participate in the open-label extension.
- HIKARI and J-RAPID studies: ACR20 non-responders at Weeks 12 or 14 were withdrawn at Week 16 and were eligible to enrol in open-label extension studies receiving CZP 200 mg Q2W. It should also be noted that the J-RAPID study included additional dosing arms of CZP (CZP 100 mg Q2W and CZP 400 mg Q2W), however, these were distinct treatment groups,

and patients did not adjust their dose of CZP during the study unless otherwise required by joining a different treatment arm due to an early escape.[6, 7] Furthermore, these alternative doses are not approved for use in the European Union.

- SWITCH study: After Week 12, all patients entered the 12 week open label extension.[4] As indicated above (response A.12), given that SWITCH is an IIS, UCB does not have access to information beyond what is provided in the manuscript.

A.18 Please clarify why a run-in period was necessary for the DOSEFLEX study but not for other trials. In addition, please clarify the length of the run-in and what changes to patients' treatment took place.

UCB response:

UCB kindly refers the ERG to Section 4.3.2 (page 57) of the UCB submission where a description of the study design and the objective of the DOSEFELX study are presented.

A.19 Please clarify why the sample size of the SWITCH trial was adjusted for potential drop-outs while the sample size of other trials was not adjusted for drop-outs.

UCB response:

For the UCB-conducted RCTs (REALISTIC, DOSEFLEX, PREDICT, J-RAPID and HIKARI), [redacted]

As per the published information, it not clear on why the investigators of the SWITCH study did adjust the sample size for potential dropouts. As per the response for A.12, given the SWITCH is an IIS, UCB does not have access to information beyond what is provided in the published manuscript and thus cannot provide further information on sample size calculations for this study.

Meta-analyses

A.20 **Priority Question**: Throughout: Please provide estimates and 95% credible intervals for between-study standard deviations where these have not been provided.

UCB response:

Between-study variance was provided in Appendix 8.12.5, page 48 of UCB's appendix submission. Credible interval for between the study variance as well as total residual deviance of the model have been provided in Table K.

Table K: Between study variance for various endpoints in the NMA

Outcome	resdev (crl)		Between study variance	Model selection
	Fixed-effects	Random-effects	Tau2 (Crl)	
ACR 20 response at 3 months	13.93 (8.037-23.60)	10.06 (3.238-20.48)	1.573 (0.02454-3.821)	Considering the limited number of studies and uncertainty in estimating the between the study variance, results from fixed-effects model should be preferred
ACR 20 response at 6 months	10.24 (3.316-21.07)	10.18 (3.269-20.76)	1.33 (0.002-3.81)	
ACR 50 response at 6 months	10.41 (3.402-21.27)	10.37 (3.416-21.01)	1.304 (0.002-3.789)	
ACR 70 response at 6 months	10.58 (3.438-21.61)	10.69 (3.499-21.72)	1.313 (0.001-3.807)	
EULAR (good) response at 6 months	6.421 (1.317-15.42)	6.38 (1.342-15.19)	1.341 (0.003-3.806)	
EULAR (good/moderate) response at 6 months	10.21 (3.334-20.93)	10.19 (3.337-20.87)	1.323 (0.002-3.798)	
Sensitivity to Combe 2012				
EULAR (good) response at 3 months	7.353 (1.827-16.74)	7.657 (2.003-17.1)	1.127 (0.00116-3.727)	
EULAR	18.1	10.11	1.737	

Outcome	resdev (crl)		Between study variance	Model selection
	Fixed-effects	Random-effects	Tau2 (Crl)	
(good/moderate) response at 3 months	(12.19-27.77)	(3.226-20.72)	(0.05919-3.845)	
EULAR (good) response at 6 months	8.536 (2.346-18.61)	8.564 (2.354-18.65)	1.34 (0.003-3.798)	
EULAR (good/moderate) response at 6 months	12.42 (4.549-24.06)	12.35 (4.557-23.84)	1.32 (0.002-3.808)	

resdev: Residual Deviation; tau2: variance of underlying distribution

A.21 Priority Question: Please clarify why approximations of results using marginal univariate normal distribution or multivariate normal distributions were preferred to using draws from the joint posterior distribution (i.e. CODA) which would have maintained correlation.

UCB response:

In the executable model submitted by UCB, the probabilities of EULAR response were modelled using posterior summaries of the baseline effect in REALISTIC (placebo and MTX group, denoted mu in NMA model), the cut-off statistic for EULAR response (denoted Z in NMA model), and treatment effect (denoted d in NMA model) from the NMA model.

By estimating the probabilities of EULAR response in the Excel model (as opposed to via the CODA output of OpenBUGS), it was possible to evaluate the influence of each parameter (reference effect, cut-off value, treatment effect) on the results of the economic analysis through one-way and scenario analyses in the Excel file. It was thus possible to determine that treatment effect was a key driver of results, and that the reference effect and cut-off statistic had limited impact on results. Using the CODA output, it would be necessary to re-run the OpenBUGS code and generate a new fixed list of CODA output for each scenario analysis, which significantly reduces the flexibility of the model.

As noted, a limitation of deriving EULAR response probabilities directly in Excel is the loss of correlation between variables. This is unlikely to significantly affect the expected probabilities of response in the model, and hence not materially impact on the expected ICER (either from the deterministic analysis or from the expectation of the probabilistic analysis) from the analysis.

A.22 **Priority Question**: Section 5.3.1.1: Please clarify how the "No response" rate for the reference treatment has been determined. As specified, "mu" is represents a study-specific parameter not a parameter for the general population.

UCB response:

As noted in response to question A21, the mean and standard deviation for the trial-specific parameter, “mu”, was obtained from the posterior summarises of the trial-specific baseline effects for the REALISTIC study, in the NMA. As such, “mu” corresponds to the probability of no response for the placebo + MTX arm of REALISTIC.

Following the approach outlined in TSD2,[9] the probability of no response is calculated as:

- $P = \Phi(\mu)$ “MTX + Placebo arm”
- $P = \Phi(\mu + d)$ “CZP + MTX arm”

Where Φ is the standard normal distribution.

In the executable model, the probabilities of response for each comparator were estimated using the baseline effect from REALISTIC, and combined with treatment effect parameters from the NMA model.

In populations A and B, the baseline effect was modelled using the placebo + MTX arm of REALISTIC, and the treatment effect parameters were estimated as bDMARD + MTX versus placebo + MTX.

In population C, the baseline effect was modelled using the CZP monotherapy arm as the baseline effect, and the treatment effect parameters were estimated as bDMARD versus CZP.

In all cases, the probabilities of EULAR response for each treatment were generated based on a population with baseline characteristics consistent with the TNFI-IR population of REALISTIC.

The REALISTIC study was chosen as the baseline effect in the economic analysis to ensure consistency between the baseline characteristics of the modelled population (i.e. baseline HAQ, pain, age, etc.), which were derived from REALISTIC, and the efficacies of treatment (i.e. EULAR response).

A.23 **Priority Question**: Table 67 (page 180 of the company submission): Please clarify where the means and standard deviations for the "No response" and cut-off came from.

UCB response:

As noted in response to A22, the mean and standard deviations for the “No response” and “cut-off” parameters were obtained from the posterior summaries of the mu and z nodes of the OpenBUGS model.

A.24 Section 4.9.2.1 (page 119 of the company submission) - Heterogeneity: Please clarify what attempt was made to explain the heterogeneity between studies and what the predictive distribution is for the effect in a new study.

UCB response:

Heterogeneity between the included studies was assessed by the I-square statistics. Please refer to Section 4.9.1.1 and Section 4.9.3 in the original UCB submission for further details on heterogeneity assessment.

For all the outcomes except EULAR good/moderate response at 3 months for the combination with MTX population, heterogeneity was observed to be low to moderate. For EULAR good/moderate response at 3 months for the combination with MTX population, I-square values were 68.9% indicating moderate-to-high heterogeneity. The heterogeneity could be by chance or could be associated with the inclusion small study effects. Considering the higher I-square value, results of random-effects meta-analysis are deemed as more reliable (Figure 42, page 121 in the original UCB submission).

Predictive distribution in a new study was calculated using STATA v11.0. Given the limited number of studies, the predictive distribution of the effect in a new study cannot be accurately estimated, with predictive interval ranging from 0 to 541270.

A.25 Section 4.9.3 (page 120 of the company submission) - Heterogeneity: Please clarify what attempt was made to explain the heterogeneity between studies and what is the predictive distribution is for the effect in a new study.

UCB response:

Please refer UCB response in A.24.

A.26 Section 4.10.3 (page 123 of the company submission): Please clarify in what sense the results of Combe 2012 are implausible and provide an explanation for why this might be the case. Please clarify why the results from Combe 2012 were deemed implausible (p123).

UCB response:

The Combe 2012 study recruited patients who had inadequate response to one or two courses of rituximab in the last 6–9 months and to at least 1 TNFi including etanercept. The sample size calculations suggested that study should recruit at least 110 patients. The study was included in sensitivity analyses and not in the base case, due to several reasons, including:

- Highly restricted inclusion criteria (inadequate response to TNFi and rituximab)

- This study was terminated due to recruitment failure. Efficacy analysis was performed on only 20 patients (10 in each group) instead of the 110 as planned.

Furthermore, the quality assessment of this study indicated a high risk of bias, and thus limited generalisability of the study results (Table 8.12.1.3, page 39-41 of the Appendices of UCB original submission).

A.27 Please clarify why Combe 2012 was not included in Figure 62 (page 189 of the company submission).

UCB response:

Combe 2012 was included as part of the sensitivity analysis to the NMA (Section 4.10.6.2, page 137 onwards of the UCB submission), but the output of this analysis was not included in the economic analysis, as the results of Combe 2012 were considered implausible (see A.26). Hence, Combe 2012 was omitted from the network diagram.

A.28 Section 4.10.5 (page 124 of the company submission) states 'An adjusted indirect analysis method was chosen over Bayesian for outcomes where evidence network included not more than two competing interventions. For evidence networks assessing more than two competing interventions a Bayesian NMA was performed'.

a. Please clarify why the distinction was made in the method of analysis and discuss what implication this has for issues such as dealing with heterogeneity and subsequent inference.

UCB response:

Adjusted indirect comparison is the first valid method available that makes it possible to perform indirect comparisons. Both adjusted indirect treatment comparison (ITC) and Bayesian NMA are equally unbiased methods. For simple networks with only two studies connected through a common comparator, an adjusted ITC was preferred over complex Bayesian NMA. For multiple competing interventions, the Bayesian NMA was preferred, since adjusted ITC would mean preceding in two by two steps due it its limitation of only considering two treatments at a time. The problem of multiple comparisons arises, leading to significance level inflation. As per French guidelines ("Indirect comparisons Methods and validity" available from www.has-sante.fr) for conducting ITC with four competing interventions, under the null hypothesis that all treatments have the same efficacy (ie. that there is no difference between treatments when considered two by two), there are still six ways of finding a difference due to chance alone (ie. six two by two comparisons). If each unit comparison is carried out with a significance level of 5%, the global risk of incorrectly finding at least one difference among the four treatments is substantially increased, rising to 26%.

Considering these limitations of adjusted ITC, Bayesian NMA was performed for networks with more than two competing interventions. Bayesian NMA is also a more suitable approach with multiple competing interventions for the following reasons:

- Flexibility of the model allowing estimation of effect sizes for multiple competing interventions in a single analysis compared to adjusted ITC where multiple two by two comparisons (significance level inflation) if ITC of more than two treatments.
- Easy to extend to handle multi-arm trials.
- Inconsistency assessment.

b. Please clarify why Bucher analyses were preferred to a Bayesian Random effects model: if possible re-analyse using a Bayesian Random Effects Model accounting for heterogeneity.

UCB response:

The Bucher analysis was conducted for the networks with two competing interventions only ie. two by two comparisons through one common comparator. Please refer to response A.28a for further details. A

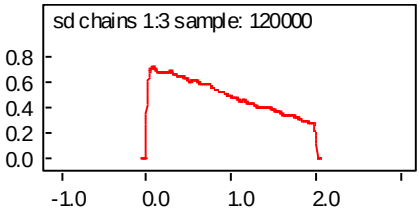
comparison of the estimates from the Bucher ITC (Section 4.10.6.1.2, page 129 of UCB's submission) and Bayesian random-effects model is presented in Table L for the ACR50 response:

Table L: Results of Bucher ITC and Bayesian Random-effects model

Outcome	Comparison	Bucher ITC, mean (95% CI)	Bayesian random-effects, mean, median (95% CrI)	Between study variance, mean, median (95% CrI)
ACR50 response 3 months	vs. TOC+MTX	0.90 (0.10-8.19)	1.77, 0.40 (0.03-8.75)	1.02, 0.59 (0.001-3.69)

The above estimates indicate that the Bayesian random-effects model is not predicting the posterior accurately, given the huge difference between mean and median values. In addition the density plots of the posterior distribution of between-study standard deviation (Figure 1) indicate that posterior is highly dominated by priors indicating little or no contribution of between study variability. Considering the uncertainty in estimating the treatment effect from Bayesian NMA, the results of Bucher ITC should be considered as more robust for very sparse networks having only two competing interventions ie. CZP and another TNFi.

Figure A: Posterior density plot for between study standard devivation-ACR50 responses 3 months



A.29 Section 4.10.5.2 (pages 126-127 of the company submission): Please clarify why there are no ACR70 data according to the following sentence whereas there are ACR20 and ACR50 data (NOTE: zeroes are classed as data): Indirect analysis was not possible for ACR70 response at 3-months due to there being no data from the comparator studies.

UCB response:

As indicated in the UCB submission (Section 4.10.5.2, page 126–127), ACR70 response data at 3-months was only available from REALISTIC and J-RAPID (both studies of CZP). In addition to these studies, ACR70 response data at 3-months was available only for tocilizumab in combination with MTX arm in the RADIATE study. However, ACR70 data was not reported for PBO in combination MTX arm and therefore this study was not considered for the ITC.

A.30 Section 4.10.6.1 (page 128-129 of the company submission): Please confirm that the results are random effects mean and not predictive distributions for the effects in new studies.

UCB response:

We confirm that the results are random-effects mean and not the predictive distributions for the effects in new studies.

A.31 Section 4.10.6.1 (page 128 of the company submission): Please clarify what is considered to be a clinically relevant non-inferiority margin for claims of "at least as effective".

UCB response:

The objective of analysis was to compare CZP versus other TNFi agents in an ITC. CZP was compared versus other comparators using an alpha error of 5% to detect a statistically significant difference. The results of the ITC demonstrated that the efficacy of CZP is comparable or superior to other comparators

(eg. EULAR good/moderate response at 6 months).Therefore it was considered that CZP was at least as effective to other comparators.

A.32 Section 5.3.1.1 (page 177 of the company submission): Please clarify what is meant by, "These parameters were varied using a series of univariate normal distributions."

UCB response:

In the probabilistic analysis for the economic analysis, the trial-specific baseline effect, treatment effect and cut-off statistic were varied using a series of univariate normal distributions (i.e. assuming no correlation between parameters). This is also described in the UCB submission, in table 67 (pages 180-181), “Effect sizes, trial-specific baseline effects, and cut-off statistics [were] varied using normal distributions in the probabilistic analysis”.

A.33 Section 5.3.1.5 (pages 189-193): Please clarify the following:

a. The expected magnitude of the between study variance given the perceived heterogeneity.

UCB response:

Since there is only one study per treatment comparison in the base case trial network (see figure 62, UCB original submission), it was not possible to estimate the “expected magnitude of between study variance” in treatment effects across the network.

b. Why a Random Effects model incorporating this was not used.

UCB response:

As stated in the UCB submission (Section 5.3.1.5, page 189), a random effects model was not considered in the base case analysis due to the limited number of studies in the network (five studies in the base case, each reporting outcomes comparing a different bDMARD versus PBO).

Issues relating to the use of random effects models for the analysis of EULAR response were also noted by the ERG in their analyses for TA375 (EULAR response in DMARD-IR). To address this, the ERG used a weakly informative prior to generate estimates of the between study variance in treatment effect in the random effects model. This analysis was based on a relatively large network of 14 studies, compared with the 5 studies available for TNFi-IR. Hence, to provide a genuine estimate of between study variance, it may be necessary to use a strongly informative prior for this parameter in the NMA model. There is however, insufficient data to inform this prior, and any misspecification of the prior may bias the outcomes of the NMA. Hence, a fixed effects model was preferred to a random effects model in the base case.

A.34 Table 69 (page 192 of the company submission): Please clarify whether the ‘Trial-specific baseline parameter’ refers to the baseline response rate for the reference treatment.

UCB response:

The trial-specific baseline parameter refers to the baseline response rate for the PBO + MTX arm of REALISTIC. In the executable model, this parameter is combined with the treatment effect parameters (denoted d in the NMA model) for bDMARD versus reference (ie. PBO control), and the cut-off statistic (denoted z in the NMA model) for EULAR response, to estimate the expected probability of response.

A.35 Section 5.3.2 (pages 193-195): Please clarify whether interaction effects were assessed.

UCB response:

No interaction effects were assessed as part of the statistical analysis.

A.36 Throughout the clinical effectiveness submission: Please provide p-values where p-values have not been calculated (for example Table 16, page 83 of the company submission) these have not been provided. Where possible, please also provide data and statistical significance for combination and monotherapy subgroups.

UCB response:

As requested, UCB have conducted additional post-hoc analyses of the outcomes reported in the original submission that previously did not include inferential statistics. These additional post-hoc results are provided in the tables below for key outcomes from the original UCB submission. It is important to highlight that all post-hoc statistical analyses are of exploratory nature (ie. are nominal p values), hence these should be interpreted with caution and no conclusion can be made on statistical significance. Furthermore, special caution should be paid to interpretation of p values generated on sample size smaller than 15% of study population.

The table below is the update to Table 16 from UCB's submission (page 83), that now includes the additional post-hoc inferential statistics (nominal p-values):

"Table 16: REALISTIC study: EULAR response rates in the TNFi experienced population (LOCF)

		PBO[a] (n=80)	CZP 200 mg Q2W[a] (n=320)	Wk 0 CZP 200 mg Q2W (OLE; n=286)[b]
EULAR Response, n (%) [OR (95% CI) p value*]				
Week 12	Good or Moderate	[redacted]	[redacted]	[redacted]
	Good	[redacted]	[redacted]	[redacted]
	Moderate	[redacted]	[redacted]	[redacted]
	None	[redacted]	[redacted]	[redacted]
Week 28	Good	[redacted]	[redacted]	[redacted]
	Moderate	[redacted]	[redacted]	[redacted]
	None	[redacted]	[redacted]	[redacted]

*CZP vs PBO
[a]FAS (LOCF); [b]Open label set (LOCF), The OLE group only includes patients randomised to CZP at Week 0, who completed 12 weeks of double-blind phase before entering the OLE. All reported p-values and confidence intervals can only be interpreted in an exploratory manner, ie, are nominal.

The table below is the update to Table 17 from UCB's submission (page 86), that now includes the additional post-hoc inferential statistics (nominal p-values):

"Table 17: REALISTIC study: DAS28(ESR) remission in TNFi experienced population during 12 week double-blind phase and OLE (NRI)

		PBO[a] (n=80)	CZP 200 mg Q2W[a] (n=320)	Wk 0 CZP 200 mg Q2W (OLE; n=286)[b]
DAS28(ESR) Remission, n (%) [OR (95% CI) p value*]				
Week 12	Remitter	[redacted]	[redacted]	[redacted]
	Non-remitter	[redacted]	[redacted]	[redacted]
Week 28	Remitter	[redacted]	[redacted]	[redacted]
	Non-remitter	[redacted]	[redacted]	[redacted]

*CZP vs PBO
[a]FAS (NRI); [b]Open label set (NRI), The OLE group only includes patients randomised to CZP at Week 0, who completed 12 weeks of double-blind phase before entering the OLE. All reported p-values and confidence intervals can only be interpreted in an exploratory manner, ie, are nominal.

The table below is the update to Table 19 from UCB's submission (page 95), that now includes the additional post-hoc inferential statistics (nominal p-values):

"Table 19: DOSEFLEX study: EULAR response rates in the TNFi experienced population (LOCF)

		CZP 200 mg Q2W + MTX* (n=42)	CZP 400 mg Q4W + MTX* (n=39)	PBO + MTX* (n=27)
EULAR Response (by Good, Moderate or None), n (%)				
Week 16†	Good	[redacted]	[redacted]	[redacted]
	Moderate	[redacted]	[redacted]	[redacted]
	None	[redacted]	[redacted]	[redacted]
Week 34	Good	[redacted]	[redacted]	[redacted]
	Moderate	[redacted]	[redacted]	[redacted]
	None	[redacted]	[redacted]	[redacted]
EULAR Response (by Response or No Response)‡, n (%) [OR (95% CI) p value*]				
Week 34	Response	[redacted]	[redacted]	[redacted]
	No response	[redacted]	[redacted]	[redacted]

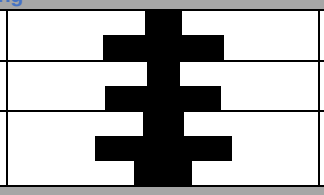
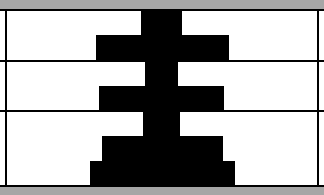
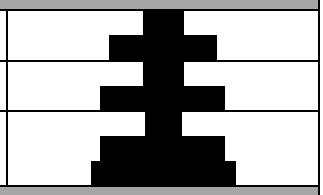
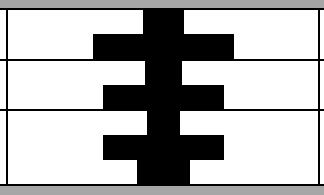
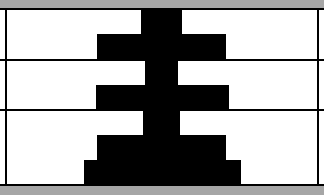
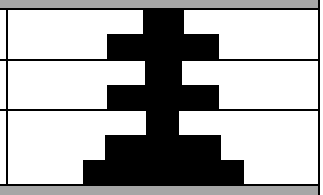
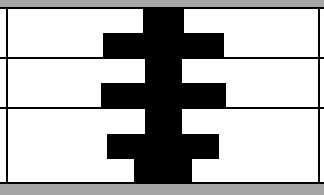
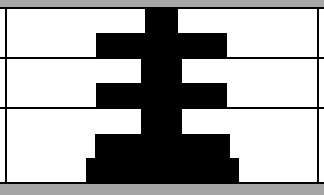
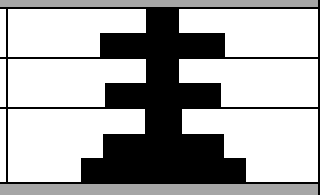
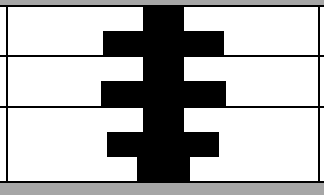
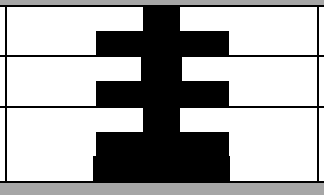
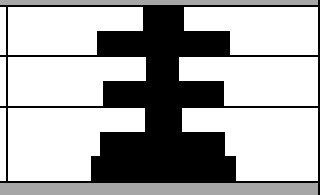
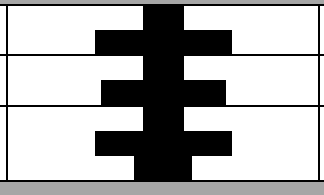
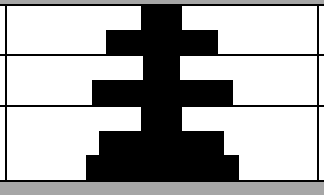
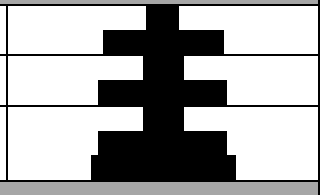
*CZP vs PBO

†All patients received CZP 200 mg Q2W during the run-in phase from Week 0–16; ‡Response defined as achieving a EULAR Response classified as Good. Non-response defined as achieving a EULAR Response classified as Moderate or No Response.

FAS (LOCF). All reported p-values and confidence intervals can only be interpreted in an exploratory manner, ie, are nominal."

The table below is the update to Table 20 from UCB's submission (page 101), that now includes the additional post-hoc inferential statistics (nominal p-values):

"Table 20: DOSEFLEX study: SF-36 domain scores for the TNFi experienced population at Weeks 0, 16 and 34 of study (LOCF)

SF-36 domain scores [n] mean (SD) {CfB p value}	PBO + MTX† (n=29)	CZP 200 mg Q2W + MTX† (n=43)	CZP 400 mg Q4W + MTX† (n=39)
Physical Functioning			
Week 0			
Week 16			
Week 34			
Role Physical			
Week 0			
Week 16			
Week 34			
Bodily Pain			
Week 0			
Week 16			
Week 34			
General Health			
Week 0			
Week 16			
Week 34			
Vitality, mean			
Week 0			
Week 16			
Week 34			
Social Functioning			
Week 0			
Week 16			
Week 34			
Role Emotional			
Week 0			
Week 16			
Week 34			
Mental Health			
Week 0			
Week 16			
Week 34			

*CZP vs PBO based on change from baseline (CfB) scores
†Final group sizes were PBO (n=28), CZP 200 mg Q2W (n=43), CZP 400 mg Q4W (n=39) at the end of Week 34. All patients received CZP 200 mg Q2W during the run-in phase from Week 0–16
FAS (LOCF). Numbers in square brackets indicate number of patients included in the analysis at each timepoint.
All reported p-values and confidence intervals can only be interpreted in an exploratory manner, ie, are nominal.

The table below is the update to Section 8.7.1.3 from the appendices of UCB's submission (page 19), that now includes the additional post-hoc inferential statistics (nominal p-values):

"8.7.1.3: REALISTIC study: EULAR response rates by TNFi experienced monotherapy and combination with MTX subgroups (LOCF)

		Monotherapy			Combination with MTX[a]		
		PBO[b] (n=23)	CZP 200 mg Q2W[b] (n=79)	Wk 0 CZP 200 mg Q2W (OLE; n=70)[c]	PBO[b] (n=51)	CZP 200 mg Q2W[b] (n=207)	Wk 0 CZP 200 mg Q2W (OLE; n=187)[c]
EULAR Response, n (%) [OR (95% CI) p value*]							
Week 12	Good or Moderate	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	Good	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	Moderate	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	None	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Week 28	Good	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	Moderate	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	None	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

*CZP vs PBO
[a]with or without other cDMARDs; [b]FAS (LOCF); [c]Open label set (LOCF), The OLE group only includes patients randomised to CZP at Week 0, who completed 12 weeks of double-blind phase before entering the OLE.
All reported p-values and confidence intervals can only be interpreted in an exploratory manner, ie, are nominal. Special caution should be paid to interpretation of p values generated on sample size smaller than 15% of study population.

The table below is the update to Section 8.7.1.4 from the appendices of UCB's submission (page 20), that now includes the additional post-hoc inferential statistics (nominal p-values):

"8.7.1.4: REALISTIC study: DAS28 scores by TNFi experienced monotherapy and combination with MTX subgroups (LOCF)

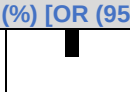
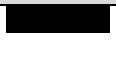
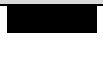
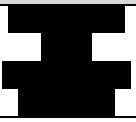
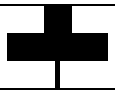
	Monotherapy			Combination with MTX[a]		
	PBO[b]	CZP 200 mg Q2W[b]	Wk 0 CZP 200 mg Q2W (OLE)[c]	PBO[b]	CZP 200 mg Q2W[b]	Wk0 CZP 200 mg Q2W (OLE)[c]
DAS28 ESR score, [n] mean (SD) {CfB, p value*}						
Week 0	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Week 12	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Week 28	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
DAS28 CRP score, [N] mean (SD)						
Week 0	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Week 12	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Week 28	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

*CZP vs PBO based on change from baseline (CfB) scores

[a]with or without other cDMARDs; [b]FAS (LOCF); [c]open label set (LOCF). The OLE group only includes patients randomised to CZP at Week 0, who completed 12 weeks of double-blind phase before entering the OLE. Numbers in square brackets indicate number of patients included in the analysis at each timepoint.
All reported p-values and confidence intervals can only be interpreted in an exploratory manner, ie, are nominal. Special caution should be paid to interpretation of p values generated on sample size smaller than 15% of study population.

The table below is the update to Section 8.7.1.5 from the appendices of UCB's submission (page 20), that now includes the additional post-hoc inferential statistics (nominal p-values):

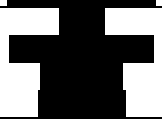
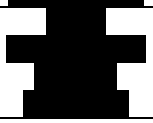
"8.7.1.5: REALISTIC study: DAS28(ESR) remission by TNFi experienced monotherapy and combination with MTX subgroups (NRI)

		Monotherapy			Combination with MTX[a]		
		PBO[b] (n=23)	CZP 200 mg Q2W[b] (n=79)	Wk 0 CZP 200 mg Q2W (OLE; n=70)[c]	PBO[b] (n=51)	CZP 200 mg Q2W[b] (n=207)	Wk 0 CZP 200 mg Q2W (OLE; n=187)[c]
DAS28(ESR) Remission, n (%) [OR (95% CI) p value*]							
Week 12	Remitter						
	Non-remitter						
Week 28	Remitter						
	Non-remitter						

*CZP vs PBO. [a]with or without other cDMARDs; [b]FAS (NRI); [c]Open label set (NRI), The OLE group only includes patients randomised to CZP at Week 0, who completed 12 weeks of double-blind phase before entering the OLE.
All reported p-values and confidence intervals can only be interpreted in an exploratory manner, ie, are nominal.[d] p value cannot be calculated due to zero cell in placebo arm. Special caution should be paid to interpretation of p values generated on sample size smaller than 15% of study population.

The table below is the update to Section 8.7.1.6 from the appendices of UCB's submission (page 21) that now includes the additional post-hoc inferential statistics (nominal p-values):

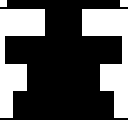
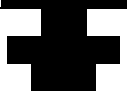
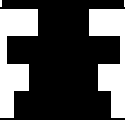
"8.7.1.6: REALISTIC study: CDAI by TNFi experienced monotherapy and combination with MTX subgroups (LOCF)

	Monotherapy			Combination with MTX[a]		
	PBO[b]	CZP 200 mg Q2W[b]	Wk 0 CZP 200 mg Q2W (OLE)[c]	PBO[b]	CZP 200 mg Q2W[b]	Wk 0 CZP 200 mg Q2W (OLE)[c]
CDAI score, [n] mean (SD) {CfB p value*}						
Week 0						
Week 12						
Week 28						

*CZP vs PBO based on change from baseline (CfB) scores
[a]with or without other cDMARDs; [b]FAS (LOCF); [c]Open label set (LOCF). The OLE group only includes patients randomised to CZP at Week 0, who completed 12 weeks of double-blind phase before entering the OLE. Numbers in square brackets indicate number of patients included in the analysis at each timepoint.
All reported p-values and confidence intervals can only be interpreted in an exploratory manner, ie, are nominal. Special caution should be paid to interpretation of p values generated on sample size smaller than 15% of study population.

The table below is the update to Section 8.7.1.7 from the appendices of UCB's submission (page 21) that now includes the additional post-hoc inferential statistics (nominal p-values):

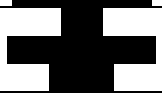
"8.7.1.7: REALISTIC study: Patient's Assessment of Fatigue by TNFi experienced monotherapy and combination with MTX subgroups (LOCF)

	Monotherapy			Combination with MTX[a]		
	PBO[b]	CZP 200 mg Q2W[b]	Wk 0 CZP 200 mg Q2W (OLE)[c]	PBO[b]	CZP 200 mg Q2W[b]	Wk0 CZP 200 mg Q2W (OLE)[c]
Patient's Assessment of Fatigue score, [n] mean (SD) {CfB p value*}						
Week 0						
Week 12						
Week 28						

*CZP vs PBO based on change from baseline (CfB) scores
[a]with or without other cDMARDs; [b]FAS (LOCF); [c]Open label set (LOCF). The OLE group only includes patients randomised to CZP at Week 0, who completed 12 weeks of double-blind phase before entering the OLE. Numbers in square brackets indicate number of patients included in the analysis at each timepoint.
All reported p-values and confidence intervals can only be interpreted in an exploratory manner, ie, are nominal. Special caution should be paid to interpretation of p values generated on sample size smaller than 15% of study population.

The table below is the update to Section 8.7.1.8 from the appendices of UCB's submission (page 21) that now includes the additional post-hoc inferential statistics (nominal p-values):

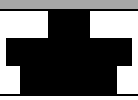
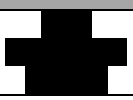
"8.7.1.8: REALISTIC study: Sleep Problem Index II score by TNFi experienced monotherapy and MTX subgroups (LOCF)

	Monotherapy			Combination with MTX [a]		
	PBO[b]	CZP 200 mg Q2W[b]	Wk 0 CZP 200 mg Q2W (OLE)[c]	PBO[b]	CZP 200 mg Q2W[b]	Wk0 CZP 200 mg Q2W (OLE)[c]
Sleep Problem Index II score, [n] mean (SD) {CfB p value*}						
Week 0						
Week 12						
Week 28						

*CZP vs PBO based on change from baseline (CfB) scores
[a]with or without other cDMARDs; [b]FAS (LOCF); [c]Open label set (LOCF). The OLE group only includes patients randomised to CZP at Week 0, who completed 12 weeks of double-blind phase before entering the OLE. Numbers in square brackets indicate number of patients included in the analysis at each timepoint.
All reported p-values and confidence intervals can only be interpreted in an exploratory manner, ie, are nominal. Special caution should be paid to interpretation of p values generated on sample size smaller than 15% of study population.

The table below is the update to Section 8.7.1.9 from the appendices of UCB's submission (page 22) that now includes the additional post-hoc inferential statistics (nominal p-values):

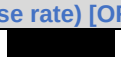
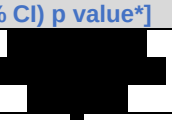
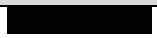
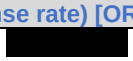
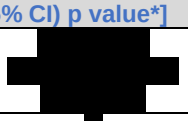
"8.7.1.9: REALISTIC study: HAQ-DI by TNFi experienced monotherapy and combination with MTX subgroups (LOCF)

	Monotherapy			Combination with MTX[a]		
	PBO[b]	CZP 200 mg Q2W[b]	Wk 0 CZP 200 mg Q2W (OLE)[c]	PBO[b]	CZP 200 mg Q2W[b]	Wk0 CZP 200 mg Q2W (OLE)[c]
HAQ-DI score, [n] mean (SD)						
Week 0						
Week 12						
Week 28						
	PBO[b]	CZP 200 mg Q2W[b]	CZP 200 mg Q2W (OLE)[c]	PBO[b]	CZP 200 mg Q2W[b]	Wk0 CZP 200 mg Q2W (OLE)[c]
HAQ-DI CfB, [n] mean (SD) p value*						
Week 12						
Week 28						

*CZP vs PBO based on change from baseline (CfB) scores
[a]with or without other cDMARDs; [b]FAS (LOCF); [c]Open label set (LOCF). The OLE group only includes patients randomised to CZP at Week 0, who completed 12 weeks of double-blind phase before entering the OLE. Numbers in square brackets indicate number of patients included in the analysis at each timepoint.
All reported p-values and confidence intervals can only be interpreted in an exploratory manner, ie, are nominal. Special caution should be paid to interpretation of p values generated on sample size smaller than 15% of study population.

The table below is the update to Section 8.7.2.1 from the appendices of UCB's submission (page 22) that now includes the additional post-hoc inferential statistics (nominal p-values):

"8.7.2.1: REALISTIC study: ACR response rates in the overall TNFi experienced population (LOCF)

	PBO (n=78)[a]	CZP 200 mg Q2W (n=317)[a]	Wk 0 CZP 200 mg Q2W (OLE) (n=285)[b]
ACR20: n (% response rate) [OR (95% CI) p value*]			
Week 12			
Week 28			
ACR50: n (% response rate) [OR (95% CI) p value*]			
Week 12			
Week 28			
ACR70: n (% response rate) [OR (95% CI) p value*]			
Week 12			
Week 28			

*CZP vs PBO. [a]FAS (LOCF); [b]open label set (LOCF); The OLE group only includes patients randomised to CZP at Week 0, who completed 12 weeks of double-blind phase before entering the OLE. Final group size at Week 28: n=275. All groups contain patients on CZP monotherapy, CZP in combination with MTX (with or without other cDMARDs) or CZP in combination with other cDMARDs (without MTX).
All reported p-values and confidence intervals can only be interpreted in an exploratory manner, ie, are nominal.

The table below is the update to Section 8.7.2.2 from the appendices of UCB's submission (page 23) that now includes the additional post-hoc inferential statistics (nominal p-values):

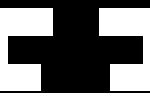
"8.7.2.2: REALISTIC study: ACR response rates by TNFi experienced monotherapy and combination with MTX subgroups (LOCF)

	Monotherapy			Combination with MTX[a]		
	PBO (n=23)[b]	CZP 200 mg Q2W (n=76)[b]	Wk 0 CZP 200 mg Q2W (OLE) (n=69)[c]	PBO (n=49)[b]	CZP 200 mg Q2W (n=207)[b]	Wk0 CZP 200 mg Q2W (OLE) (n=187)[c]
ACR20: n (% response rate)						
Week 12	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Week 28	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
ACR50: n (% response rate)						
Week 12	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Week 28	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
ACR70: n (% response rate)						
Week 12	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Week 28	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

*CZP vs PBO. [a]with or without other cDMARDs; [b]FAS (LOCF); [c]Open label set (LOCF); The OLE group only includes patients randomised to CZP at Week 0, who completed 12 weeks of double-blind phase before entering the OLE. Final groups size at Week 28: n=68 and n=179, for monotherapy and combination with MTX, respectively. All reported p-values and confidence intervals can only be interpreted in an exploratory manner, ie, are nominal. Special caution should be paid to interpretation of p values generated on sample size smaller than 15% of study population.

The table below is the update to Section 8.7.3.1 from the appendices of UCB's submission (page 24) that now includes the additional post-hoc inferential statistics (nominal p-values):

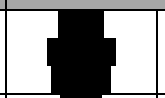
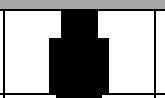
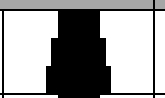
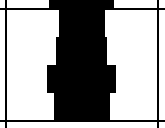
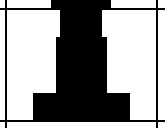
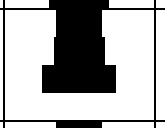
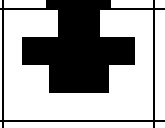
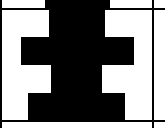
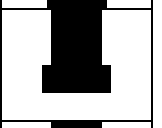
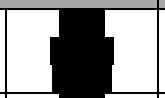
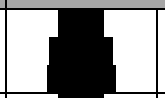
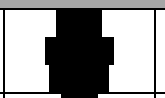
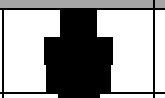
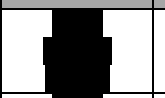
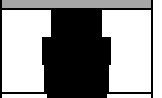
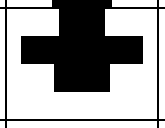
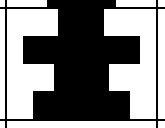
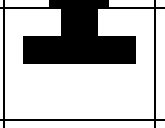
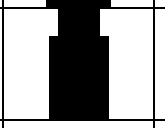
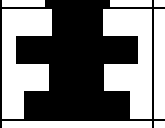
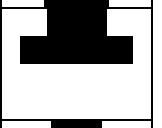
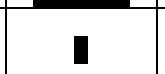
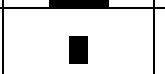
"8.7.3.1: REALISTIC study: ACR component scores in the TNFi experienced population (LOCF)

ACR component score, [n] mean (SD) {CfB p value*}	PBO[a]	CZP 200 mg Q2W[a]	Wk 0 CZP 200 mg Q2W (OLE)[b]
SJC28			
Baseline			
Week 12			
Week 28			
TJC28			
Baseline			
Week 12			
Week 28			
Pain (PtAAP)			
Baseline			
Week 12			
Week 28			
PtGADA			
Baseline			
Week 12			
Week 28			
PhGADA			
Baseline			
Week 12			
Week 28			

*CZP vs PBO based on change from baseline (CfB) scores
[a]FAS (LOCF); [b]Open label set (LOCF). The OLE group only includes patients randomised to CZP at Week 0, who completed 12 weeks of double-blind phase before entering the OLE. Numbers in square brackets indicate number of patients included in the analysis at each timepoint.
All reported p-values and confidence intervals can only be interpreted in an exploratory manner, ie, are nominal.

The table below is the update to Section 8.7.3.2 from the appendices of UCB's submission (page 25) that now includes the additional post-hoc inferential statistics (nominal p-values):

"8.7.3.2: REALISTIC study: ACR component scores by TNFi experienced monotherapy and combination with MTX subgroups (LOCF)

	Monotherapy			Combination with MTX[a]		
ACR component score, [n] mean (SD) {CfB p value*}	PBO[b]	CZP 200 mg Q2W[b]	Wk 0 CZP 200 mg Q2W (OLE)[c]	PBO[b]	CZP 200 mg Q2W[b]	Wk 0 CZP 200 mg Q2W (OLE)[c]
SJC28						
Baseline						
Week 12						
Week 28						
TJC28						
Baseline						
Week 12						
Week 28						
Pain (PtAAP)						
Baseline						
Week 12						
Week 28						
PtGADA						
Baseline						
Week 12						
Week 28						
PhGADA						
Baseline						
Week 12						
Week 28						

*CZP vs PBO based on change from baseline (CfB) scores
[a]with or without other cDMARDs; [b]FAS (LOCF); [c]Open label set (LOCF). The OLE group only includes patients randomised to CZP at Week 0, who completed 12 weeks of double-blind phase before entering the OLE. Numbers in square brackets indicate number of patients included in the analysis at each timepoint.
All reported p-values and confidence intervals can only be interpreted in an exploratory manner, ie, are nominal. Special caution should be paid to interpretation of p values generated on sample size smaller than 15% of study population.

The table below is the update to Section 8.8.2.1 from the appendices of UCB's submission (page 27) that now includes the additional post-hoc inferential statistics (nominal p-values):

"8.8.2.1: DOSEFLEX study: ACR response rates in the TNFi experienced population (LOCF)

	PBO + MTX (n=28)	CZP 200 mg Q2W + MTX (n=43)	CZP 400 mg Q4W + MTX (n=39)
ACR20: n (% response rate) [OR (95% CI) p value*]			
Week 16†	[redacted]	[redacted]	[redacted]
Week 34	[redacted]	[redacted]	[redacted]
ACR50: n (% response rate) [OR (95% CI) p value*]			
Week 16†	[redacted]	[redacted]	[redacted]
Week 34	[redacted]	[redacted]	[redacted]
ACR70: n (% response rate) [OR (95% CI) p value*]			
Week 16†	[redacted]	[redacted]	[redacted]
Week 34	[redacted]	[redacted]	[redacted]

*CZP vs PBO
†Open-label run-in phase: all groups, including PBO, received CZP in the 16-week open-label run-in phase, FAS, LOCF. All reported p-values and confidence intervals can only be interpreted in an exploratory manner, ie, are nominal."

The table below is the update to Section 8.8.3.1 from the appendices of UCB's submission (page 28) that now includes the additional post-hoc inferential statistics (nominal p-values):

"8.8.3.1: DOSEFLEX study: ACR component scores in the TNFi experienced population (LOCF)

ACR component score, [n] mean (SD) {CfB p value*}	CZP 200 mg Q2W + MTX*	CZP 400 mg Q4W + MTX*	PBO + MTX*
SJC28			
Baseline			
Week 16†			
Week 34			
TJC28			
Baseline			
Week 16†			
Week 34			
Pain (PtAAP)			
Baseline			
Week 16†			
Week 34			
PtGADA			
Baseline			
Week 16†			
Week 34			
PhGADA			
Baseline			
Week 16†			
Week 34			

*CZP vs PBO based on change from baseline (CfB) scores
†All patients received CZP 200 mg Q2W during the run-in phase from Week 0–16
FAS (LOCF). Numbers in square brackets indicate number of patients included in the analysis at each timepoint.
All reported p-values and confidence intervals can only be interpreted in an exploratory manner, ie, are nominal.

B. Clarification on cost-effectiveness data

B.1 **Priority Question**: It is noted that other bDMARDs are only recommended for severe patients (defined as a DAS28 score > 5.1) after the failure of a TNF inhibitor (TNFi). It is further noted that the company submission does not include an assessment of the clinical and cost effectiveness of certolizumab pegol in moderate to severe RA. Please clarify whether the ICERs presented should therefore be assumed to be applicable to severe patients only, particularly given that the patient characteristics in the model are based on the TNFi experienced subgroup of REALISTIC, which has a mean DAS28 score of [redacted]. If this is not the case please provide an ICER for the moderate to severe RA population against the appropriate comparator.

UCB response:

As per the original submission, UCB would like to clarify that the ICERs presented in the submission are applicable to patients with moderate to severe RA as per the scope of the appraisal, and described in Section 5.2.1 of UCB's submission. As noted, the patient characteristics in the model are based on the REALISTIC study, which enrolled a mixed population of patients with moderate (n=26, TNFi experienced) and severe RA (n=371, TNFi experienced).

B.2 **Priority Question**: Please clarify why certolizumab pegol was considered for population A as an addition to the treatment sequence before rituximab instead of as an alternative to rituximab. It is noted that this elongates the sequence. Please clarify why, if elongated sequences are being considered certolizumab pegol was not positioned after rituximab and an incremental analysis performed between these two strategies.

UCB response:

The sequence of therapies considered for population A were selected based on the consultation with an external clinical rheumatologist, with experience in treating RA patients in clinical practice in England. The clinical expert opinion was that it would be clinically reasonable to consider CZP before rituximab, unless contraindicated, to allow a second TNFi treatment option, before switching to another mechanism of action agent. Following the existing NICE recommendations, rituximab therapy would thus be followed by tocilizumab and other standard therapies in both sequences.

An incremental analysis comparing CZP before rituximab versus after rituximab was not deemed relevant for inclusion and thus not performed for a number of reasons:

- The decision as to whether CZP can be given as an alternative to therapies in patients withdrawn from rituximab or MTX, is addressed separately in populations B and C of the economic analysis;

- The submitted model was designed to assess the cost-effectiveness of CZP at each point in the sequence, as outlined in the final scope of the appraisal. The model structure considers a number of structural simplifications surrounding the efficacy of subsequent therapies (ie. constant probabilities of discontinuation), given limitations in the evidence surrounding the efficacy of subsequent therapies, such as conventional DMARDs (cDMARDs) given after inadequate response to bDMARDs. Hence, the submitted model is not designed to simultaneously assess the optimal positioning of CZP in the RA treatment pathway, as this was considered not to be in line with the final scope of this appraisal.

B.3 **Priority Question**: Please clarify why abatacept was included after tocilizumab in the treatment sequence for population A (Table 64, page 174 of the company submission). This appears to be outside of NICE's recommendations for abatacept.

UCB response:

The sequence of therapies considered for population A were selected based on consultation with an external clinical rheumatologist, with experience in treating RA patients in clinical practice in England. The clinical expert feedback was that after failure of TNFi, patients would likely receive a sequence containing rituximab, tocilizumab and abatacept, if the patient remains tolerant to MTX.

The current NICE recommendations on the use of abatacept after failure on TNFi state that:

"*Adalimumab, etanercept, infliximab and abatacept, each in combination with methotrexate, are recommended as treatment options only for adults with severe active rheumatoid arthritis who have had an inadequate response to, or have an intolerance of, other DMARDs, including at least one TNFi,* ***and who cannot receive rituximab therapy because they have a contraindication to rituximab, or when rituximab is withdrawn because of an adverse event****.*"

Therefore the decision to position abatacept after tocilizumab in population A was justified on the basis that:

- NICE recommendations permit the use of abatacept in patients who develop a contraindication to rituximab or are withdrawn from therapy because of an AE. All patients in the sequence would have received rituximab prior to tocilizumab, and hence may be eligible for therapy as per this recommendation.

- The recommendation refers to “at least one” TNFi, which supports its use after two or more TNFis, including CZP in this population.

- External clinical advice that in TNFi-IRs, abatacept would be provided after failure on rituximab and tocilizumab.

B.4 **Priority Question**: Please provide a scenario analysis where mortality is not affected by changes in the HAQ score and is only dependent on baseline HAQ, as per the Assessment Group’s assumptions in NICE technology appraisal 375 ‘Adalimumab, etanercept, infliximab, certolizumab pegol, golimumab, tocilizumab and abatacept for rheumatoid arthritis not previously treated with DMARDs or after conventional DMARDs only have failed’.

UCB response:

The results of the requested sensitivity analysis are presented in Table M (population A), Table N (population B), and Table O (population C).

The analysis was conducted in two steps. First, the mortality adjustment for HAQ was set to 1 in cell “MORT_MULT_DSA” of the mortality sheet. This removes any adjustments for HAQ over time. Secondly, the natural history mortality rates used in the model were inflated in line with the baseline HAQ (1.55) and the mortality multiplier of 1.43 per unit change in HAQ. The estimated multiplier for all-cause mortality was 1.87.

The results of the requested sensitivity analysis were found to be consistent with the expected impact of assuming no change in mortality with worsening HAQ-DI. That is, the sensitivity analysis predicted:

- Longer life expectancy compared to the base case, when assuming mortality is dependent on baseline HAQ-DI (1.55) alone, and independent of the changing HAQ-DI status of the population (1.55 raising to 2.76 over time)

- By removing the effect of changing HAQ-DI on mortality, the model predicts no difference in life years across all therapies

In these sensitivity analyses, CZP was the optimal treatment strategy at conventional thresholds of £20,000 to £30,000 per QALY gained in population B and C, and was cost-effective versus rituximab at a threshold of approximately £34k in population A.

Across all populations, the ICERs presented in the scenario analysis were consistent with the results and conclusions of the base case analysis submitted by UCB, and hence show the robustness of the model to assumptions on the relationship between HAQ-DI and mortality.

Table M: Results of sensitivity analysis assuming mortality is associated with baseline HAQ-DI (Population A)

Technologies (branded biosimilar)	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER (£/QALY) versus baseline	ICER (£/QALY) incremental
RTX + MTX (MabThera)	£140,237	16.266	6.946					Cost effective at WTP < £34,306
CZP+ MTX (Cimzia)	£149,165	16.266	7.206	£8,927	0.000	0.260	£34,306	**Cost effective at WTP > £34,306**
Deterministic results; ordered in terms of least to most expensive; ICER, incremental cost effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life years;								

Table N: Results of sensitivity analysis assuming mortality is associated with baseline HAQ-DI (Population B)

Technologies	Total costs (£)	Total LYG	Total QALYs	Incremental costs versus baseline (£)	Incremental LYG versus baseline	Incremental QALYs versus baseline	ICER (£/QALY) versus baseline	ICER (£/QALY) incremental
GOL + MTX (Simponi)	£100,480	16.266	6.050					**Cost effective at WTP < £2,551**
ADA + MTX (Humira)	£100,480	16.266	6.050	£0	0.000	0.000	-	
ETA + MTX (Enbrel)	£100,900	16.266	6.050	£420	0.000	0.000	Dominated	Dominated
CZP + MTX (Cimzia)	£101,136	16.266	6.307	£656	0.000	0.257	£2,551	**Cost effective at WTP between £2,551 and £133,744**
IFX + MTX (Remicade)	£104,761	16.266	6.050	£4,281	0.000	0.000	Dominated	Dominated
ABA + MTX (IV - no PAS) (Orencia)	£118,661	16.266	6.097	£18,181	0.000	0.046	£391,521	Dominated
TOC + MTX (IV - no PAS) (RoActemra)	£127,217	16.266	6.502	£26,737	0.000	0.452	£59,118	**Cost effective at WTP > £133,744**
Deterministic results; ordered in terms of least to most expensive; ICER, incremental cost effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life years								

Table O: Results of sensitivity analysis assuming mortality is associated with baseline HAQ-DI (Population C)

Technologies	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER (£/QALY) versus baseline	ICER (£/QALY) incremental
ADA (Humira)	£99,135	16.266	5.885					**Cost effective at WTP < £3,798**
ETA (Enbrel)	£99,536	16.266	5.885	£401	0.000	0.000	Dominated	Dominated
CZP (Cimzia)	£100,116	16.266	6.143	£981	0.000	0.258	£3,798	**Cost effective at WTP between £3,798 and £127,949**
TOC (IV – no PAS) (RoActemra)	£126,032	16.266	6.345	£26,897	0.000	0.461	£58,379	**Cost effective at WTP > £127,949**
Deterministic results; ordered in terms of least to most expensive; ICER, incremental cost effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life years								

B.5 **Priority Question**: The values for changes in HAQ score from RADIATE (-0.39 for biologics, -0.05 for methotrexate) are presumed to be values for the entire population. It is anticipated that the HAQ improvements for good and moderate responders would be greater than for the entire population. Please clarify whether this has been considered, noting that 32.3% and 83.5% of patients were non-responders in the tocilizumab and methotrexate groups respectively. In addition, please clarify why no analyses were run assuming the same HAQ change based on EULAR response independent of treatment.

UCB response:

As noted in the submission document (Section 5.3.5, page 199), "the efficacy of subsequent biologics and cDMARDs were modelled using data from the RADIATE study, where approximately 50% of the enrolled population had received two or more TNFis prior to baseline". As noted by the ERG, the values used in the model correspond to the change in HAQ from the entire population of RADIATE. Data on the change in HAQ by responder status, or by number of prior TNFis was not reported in RADIATE, and hence the full population values were used.

No sensitivity analyses were performed on this parameter, as the short-term efficacies of subsequent therapies were not considered significant drivers of the results of the evaluation, relative to other clinical parameters, such as the efficacy of first therapy in the sequence. It is however acknowledged that there is limited data on the efficacy of therapy in a heavily pre-treated population (two/three plus prior biologics), and this is why a simplified approach to modelling the efficacy of subsequent therapies was taken.

As noted by the ERG, no analyses were performed assuming the same HAQ-DI change based on EULAR response independent of treatment. This analysis requires the assumption that the proportion of responders (moderate or good) who are good responders is the same between bDMARD and cDMARD treated patients. This assumption does not appear to be supported by data from ATTAIN, REFLEX, and REALISTIC[2] (see Table P; new data in support of this response), where the proportion of good versus moderate responders was consistently greater in the bDMARD treated population than in the cDMARD control groups[3]. As the good EULAR response has been shown to be associated with a greater improvement in HAQ-DI than the moderate EULAR response, it is expected that patients continuing on bDMARDs will experience a greater improvement in HAQ-DI, versus those continuing on cDMARDs. Hence, it was not considered appropriate to assume the same HAQ-DI change for both bDMARDs and cDMARD treatments.

[2] RADIATE and GO-AFTER only reported the proportion of moderate or good responders.

[3] The proportion of responders (moderate or good) who were good responders were estimated from available data from ATTAIN, REFLEX, and REALISTIC. The ratio of proportion of responders (moderate or good) who were good responders was calculated comparing active versus control arms; A ratio greater than one would indicate a higher proportion of good responders in the active arm.

Table P: Summary of data showing the proportion of moderate or good responders who are good responders in the active and control arms of available clinical trials

Study	Comparison	PBO control arm		Active arm		Comparison
		None/ moderate/ good N	% of moderate or good responders who are good responders	None/ moderate/ good N	% of moderate or good responders who are good responders	Ratio of % of moderate or good responders who are good responders (active vs PBO)
ATTAIN (6 month)	Abatacept + cDMARD versus PBO + cDMARD	100/29/4	12.1%	128/100/30	23.1%	1.91
REFLEX (6 month)	Rituximab + MTX versus PBO + MTX	165/40/4	9.1%	117/149/45	23.2%	2.55
REALISTIC (3 month)	CZP + MTX versus PBO + MTX	26/19/6	24.0%	50/84/53	38.7%	1.61

B.6 Please clarify why certolizumab pegol was not considered for population A instead of tocilizumab in patients with an inadequate response to rituximab.

UCB response:

In line with the scope for the appraisal, the comparator sequence for population A consisted of rituximab followed by conventional standard therapies that are currently routinely used in the NHS. This includes the use of tocilizumab after rituximab, the only bDMARD treatment currently recommended by NICE in this line of therapy. Since CZP is not currently recommended after rituximab, then its consideration in the treatment sequence for population A would be outside of NICEs recommendations for CZP. A similar issue is highlighted by the ERG, in relation to abatacept in question B3, above.

B.7 Please clarify why non-biologics were included as a combined therapy for population A and as individual lines of therapy for population B and C.

UCB response:

Where feasible, a set sequence of follow-on therapies (A followed by B, followed by C, etc.) was considered for all populations in the economic analysis. This is because the efficacy of a combined non-bDMARD group would have to be assumed, given that no study has reported outcomes for the combined non-biologic group assumed in the model.

In population A, the number of lines of therapy exceeded the maximum number permitted in the model (maximum of eight lines), when considering both elongated sequences and a set sequence of follow-on therapies. Hence, in this population, it was necessary to consider a combined non-biologics group in the modelled pathway. In populations B and C, a set sequence of follow-on therapies was used.

B.8 Please clarify why a simple linear regression was used to model the relationship between HAQ-DI and pain (Figure 64, page 195 of the company submission) instead of a quadratic model as in NICE technology appraisal 375.

UCB response:

A simple linear regression model was used to model the relationship between HAQ-DI and pain, on the basis that:

- Pain and HAQ-DI scores were found to be approximately linear across the range of mean HAQ-DI scores permitted in the model base case (0 to 2.76). From visual inspection of the curve, the non-linear relationship between HAQ-DI and pain appears most prominent at HAQ-DI scores greater than 2.76, which are outside the bounds of the base case analysis

- The cohort-based Markov model cannot easily accommodate non-linear terms for predictive equations (ax^2) (see response to B.11)

B.9 Please clarify why for the discontinuation of second and subsequent treatments a exponential distribution was used instead of the Weibull distribution used for the first treatment. Please provide the justification as to why the transition probability for second and subsequent treatments were assumed to equal that for the Weibull distribution between months 6 and 12.

UCB response:

The time spent on second or follow-up treatment is modelled using a series of health states that track the treatment status of the cohort. For each line of follow-up therapy (second, third, fourth, etc), the model includes two health states representing the first six months and all subsequent periods of follow-up treatment. Through these states, it is possible to estimate the number of patients who have started therapy in a given cycle and occupying the "first six month" state, and the number of patients who are continuing on their therapy from the previous cycle and occupying the "all subsequent period" state. This makes it possible to:

- Allow costs to vary between the first and subsequent six months of follow-up therapy, if required

- To account for differences in the probability of continuing follow-up therapy after the first six months, where initial response to follow-up therapy is assessed, and all subsequent six month periods, where durability of response is assessed

- Appropriately account for the short-term effects of treatment, in terms of initial gains in HAQ, and the long-term effects of continued therapy on HAQ score (i.e. deteriorating HAQ over time for patients treated on cDMARDs)

These aspects are considered important drivers of costs and health effects in an RA population treated with a sequence of therapies.

As noted in the submission, the probability of continuing follow-up (ie second and subsequent) therapy after the first six months of treatment was modelled using EULAR response probabilities from the NMA (all follow-up therapies including rituximab, tocilizumab, abatacept and cDMARDs). This is in line with UK clinical practice. All patients who respond to and survive on follow-up therapy are assumed to enter the all subsequent period state. In this state, the probability of discontinuing therapy during all subsequent six monthly cycles is modelled assuming a constant transition probability. The probability of discontinuing follow-up therapy during all subsequent six month periods was modelled using a constant transition probability because once a patient enters the "all subsequent period" state their history in terms of time spent on subsequent therapy is lost. Without this information, it is not feasible to appropriately allocate the time-varying transition probabilities generated from a Weibull distribution to the modelled cohort. Hence the exponential distribution was used to model the probability of discontinuing follow-up therapy instead of Weibull.

For bDMARDs, the constant transition probability was set equal to the probability of discontinuing therapy between months 6 and 12 of the BSRBR curve. This ensured that the probability of discontinuing therapy in all subsequent periods was conditional on having continued therapy after the first six month of treatment.

In order to continuously track the history of the cohort, it would be necessary to further divide the "all subsequent period" state into a series of sub-states representing each six month period from six months to time horizon, which in turn, would significantly increase the number of states in the model. This would have added little additional accuracy in modelling outcomes, given limited evidence supporting

outcomes of second or subsequent follow-up treatment in TNF-IR patients and would have overly complicated the programming of the model.

B.10 Please clarify why the mapping between HAQ and EQ-5D from Brennan et al. was used in the base case analysis instead of the more recent one used within the Assessment Group model for NICE technology appraisal 375 (i.e. Hernandez-Alava et al.).

UCB response:

In the submitted base case, HAQ-DI scores are mapped to EQ-5D utilities using the mapping algorithm from Brennan et al, which reported a mean change in EQ-5D per unit change in HAQ of -0.2012. Of note, these regressions were based on results from a study by Bansback et al 2007, initially published at the end of Brennan et al. 2006. These regressions have been used by Brennan and colleagues in a study to evaluate the cost effectiveness of TNF inhibitors over conventional DMARD therapy in RA. Bansback et al. examined patients with RA participating in 2 studies in the UK (n = 151) and Canada (n = 319) who completed the HAQ, EQ-5D, and SF-36. Models were developed of the relationship between the HAQ-DI and SF-6D and EQ-5D using regression analyses. The mapping regression from Brennan et al was identified through a review of existing algorithms deriving utility weights from HAQ-Di scores, used to inform the previous submitted cost effectiveness assessment of CZP in NICE TA375.

As indicated in the original UCB submission, the regression coefficients used by the Assessment Group during the NICE TA375 were tested in a sensitivity analysis, based on the finite mixture models presented by Hernandez Alava et al, and which explore the relationship between EQ-5D, HAQ and pain. The full equations for Hernandez Alava et al were not implemented in the cohort model to avoid over complicating the model programming. Instead, the coefficients for HAQ (linear term) and pain for each of the four classes of Hernandez Alava et al were programmed in the model with the option of selecting any one of the classes, as required. All other parameters, including non-linear terms for HAQ were excluded from the model. A series of sensitivity analyses were then performed using the covariates from each of the regression classes, to provide a range of potential outcomes based on Hernandez Alava et al.

B.11 Please clarify why the implementation of Hernandez-Alava et al. utility mappings only includes a subset of the covariates, and why the distributions for each of the classes of the mixture model are independently applied instead of combined through a weighted average (using the class membership probabilities as weights). In addition, please clarify to what extent inaccuracy within the calculation of utility values would impact on the resultant ICERs.

UCB response:

Only the linear terms for the Hernandez-Alava et al mapping algorithms were implemented in the submitted model, for the following reasons (Table Q):

- At the time of development, there was no clear approach to incorporating the non-linear terms (i.e.HAQ^2) of the Hernandez-Alava et al mapping algorithm in a cohort-based model that calculates outcomes on expected values. In particular, as shown below, it was found that the expectation of HAQ^2 cannot be directly inferred from the expectation of HAQ (see example below, where E[HAQ^2] does not equal (E[HAQ])^2)). At a cohort-level, it is therefore not clear how changes in the expectation of HAQ would impact on the expectation of HAQ^2. To avoid generating erroneous results, all non-linear terms were dropped from the model.

Table Q: Example showing HAQ and HAQ^2 scores at patient level, and values calculated through expected values

Pat.Id	HAQ	HAQ2
1	1.5	2.25
2	1.2	1.44
3	1.7	2.89
E[HAQ]	1.46667	-

E[HAQ^2]	-	2.19333
E[HAQ]^2	2.15111	

- Each class of Hernendez et al, was applied independently in the model to provide a range of plausible ICER's in the sensitivity analysis. The class membership functions that are used to weight each class, were not implemented in the model as it is only possible to estimate the probability of class membership conditional on the expected baseline characteristics of the cohort, which will differ to the expected probability of class membership conditional on patient-level baseline characteristics, which is the preferred input for the cohort model. Similar to the above, it was not clear whether this discrepancy would significantly bias the results of the analysis, and hence they were omitted from the model.

By implementing each class independently, it is possible to see a potential "spread" of plausible ICERs that may represent the range of outcomes possible with Hernandez et al. It is expected that the introduction of the weights and non-linear terms, if feasible, to the model would yield an ICER in the range of values presented in the sensitivity analysis submitted by UCB.

B.12 Please clarify why a Dirichlet distribution was used for the weight parameter for the PSA rather than fitting a distribution. This may have an impact should the number of vials required change within a weight band.

UCB response:

Following approaches adopted in previous NICE appraisals for CZP in RA, the weight distribution of the population was defined by a series of weight classes, ranging from 0-39 kg to 120–200 kg. Between 40–120 kg, the weight class boundaries are spaced 5 kg apart, ie. 40–44.9 kg. All patients in the submitted model were assigned to one of the weight class boundaries.

Following standard practice in the literature, a multinomial beta (or Dirichlet) distribution was used to sample the proportion of patients in each weight class in the model. The summation of the weight classes was equal to 100% throughout the PSA. The Dirichlet was preferred to fitting a distribution directly to the data as the resulting distribution of weight class membership is not constrained to a single shape or form (ie. single peak and symmetrical if predicted by a truncated normal).

B.13 Please clarify why the cycle length is different (3 months instead of 6 months) for the second and third cycles of the model. Please confirm whether there is an error in cycles 2 and 3 with the 6-month discontinuation probabilities being applied to 3 monthly time cycles.

UCB response:

The submitted economic model includes cycle lengths of three instead of six months for the second and third cycles. The probability of discontinuation and mortality for first therapy are adjusted for the shorter cycle length, as shown below:

- **Probability of discontinuation**: Survival calculations sheet rows 9 to 10 (Probability of discontinuation in cycles two and three calculated as 1-S(0.75)/S(0.5) and 1-S(1.0)/S(0.75))

- **Hazard rate of death/mortality**: Survival calculations sheet rows 103 to 104 (values adjusted to 0.5 * six month hazard rate)

The QALY calculations are also adjusted for the shorter cycle periods.

There are however some discrepancy in the calculation of the probability of discontinuation for follow-up therapies, costs and HAQ change, that have been corrected. A more detailed explanation of the revisions to the executable model is provided in Appendix 3. The updates to the model were checked by an independent modeller, and extreme value tests were performed to ensure the model produced consistent and logical results.

The basecase results of the corrected model (Appendix 3) are consistent with the original base case submitted, such that the discrepancy in the above calculations had limited impact on the estimated ICERs, and had no impact on the conclusions of the base case analysis, which demonstrates that CZP is a cost-effective therapy in TNF-IR.

The revised deterministic and probabilistic results for the base case analysis are included in Appendix 3. It is expected that the results of the sensitivity analyses will be similar to those presented in the original UCB submission, given that there were only minor differences in base case results between the original and revised models (Appendix 3, Table 1).

B.14 Please clarify why nurse visits for a percentage of patients unwilling or unable to perform subcutaneous injections were not considered as in the Assessment Group's economic model for NICE technology appraisal 375.

UCB response:

Based on consultation with an external clinical rheumatologist, experienced in treating RA patients in clinical practice in England, it was assumed that help for patients unwilling or unable to perform subcutaneous injections was provided through the home help scheme funded by the manufacturer. Thus, it was assumed that any costs associated with supporting the administration of subcutaneous injections were not passed to the health service.

B.15 Please clarify the following:

a. Why the half-cycle correction implementation counts the utility of the first cycle 1.5 times instead of 0.5 times as it is customary when applying the half-cycle correction (or just once, if the first cycle should be exempt of half-cycle correction).

UCB response:

Following standard Pharmacoeconomics guidelines, a lifetable adjustment approach was adopted when calculating the total costs and health effects of treatment in the model. The primary purpose of this adjustment is to correct for potential biases from not knowing the exact time of transition into a given state. In the lifetable approach, the "average" occupancy of the state is taken based on the numbers occupying the state at the beginning and end of the cycle.

This is problematic in an area such as RA, where treatment rules mean that patients may simultaneously enter one state at a fixed time point but gradually transition into others. Examples include the assessment of response to first therapy, where all non-responders are assumed to instantly and simultaneously transition to the next subsequent therapy state at the end of the first cycle. After the first cycle, patients who discontinue first therapy may discontinue at any point during the cycle. Hence, it was necessary to construct an adapted set of lifetable calculations to accommodate both the instantaneous and gradual transitions in the cohort lifetime.

A detailed explanation of the calculation of utilities in the first cycle is shown below, and in response to questions b) and c).

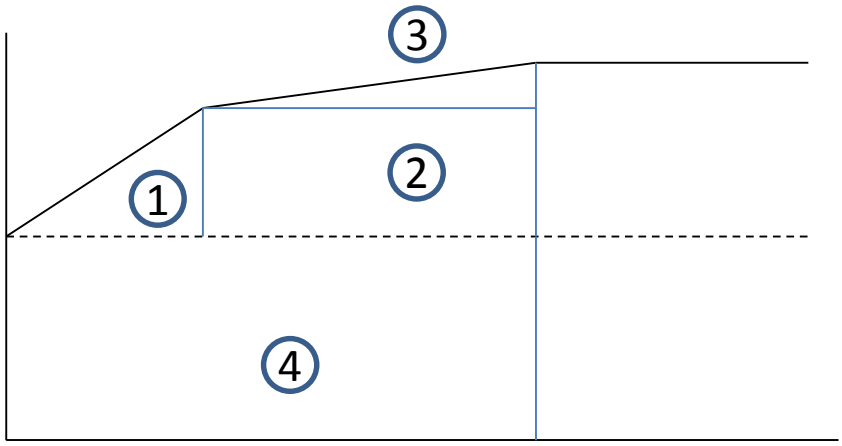
The calculation of health state utilities is presented in columns GB through GN on each of the Markov trace sheets. The first cycle utility calculations are outlined in cells GB27 through GE27.

In the first cycle of the model, the QALY assigned to the population is modelled based on an area under the curve type analysis, which assumes the following:

- 100% of the utility gain from the response is achieved by week 6 of the first cycle period
- The utility gain is maintained from week 6 to 6 months

This area under the curve calculation is composed of four segments as shown below:

Figure C: Calculation of the quality adjusted life years gained for the first cycle of the model



1 (Change in Utility x (X%) x LY (6/52))/ 2

2 Change in Utility x (X%) x LY ((26-6)/52)

3 (Change in Utility x (1-X%) x LY ((26-6)/52)/ 2

4 Baseline utility x Time (26/52)

The corresponding Excel equation for the QALYs associated with the first cycle follow this equation:

=
K27*(((GB25*GC13*GC16)/2+(GB25*GC13*GC17)+(GB25*GC14*GC17)/2)+(MP_BASEEQ5D)*($GA27))

Or/

= number of patients in response state * (((change in utility * 100% * 6/52)/2 + (change in utility *100%*20/52) + (change in utility *0%*20/52)/2))+(baseline EQ-5D) * (0.5))

To validate the calculation of QALYs for the first cycle, the following extreme value tests were performed:

- Set baseline utility to zero and change in utility to zero (correctly predicts 0 QALY)
- Set baseline utility to one, change in utility to zero, and mortality to zero (correctly predicts 0.5 QALY)

The model performed as expected, and generated logical results.

b. Why in some cases (e.g. Markov_C!CK28) the half cycle correction mixes utility values from two different states.

UCB response:

In the example of Markov_C!CK28, the calculation (K27 + (W28 – K27)*0.5) can be explained as follows:

- As mentioned in response to Question B15 a, K27 corresponds to the number of patients classified as inadequate responders at the end of the first cycle/ start of the second cycle. These patients are assumed to instantly and simultaneously transition to the first subsequent therapy – six month state, at the end of the responder phase. Hence, these patients are assumed to occupy this state for the full cycle period of the model, and contribute K27 * 1 to the average state occupancy calculation

- (W28-K27) corresponds to the number of patients in the state at the end of the cycle (W28) who are not inadequate responders at the start (K27). These patients comprise those who responded at six months, but discontinued by the end of the following cycle. None of the "discontinuing responder patients" had occupied the state at the start of the cycle, and so the "average" state occupancy for this population is (0+(W28-K27))/2 = (W28-K27)*0.5

To note, the calculation (K27 + (W28-K27)*0.5) also simplifies to (K27 + W28)/2, i.e. the average occupancy assuming all patients enter the state.

All other states in the first cycle period are subject to the usual lifetable correction (i.e. [number in state at start of cycle + number in state at end of cycle]/2). The total number of patients in the "lifetable" adjusted population is calculated in row CZ. A standard check comparing the total number of patients in the model versus the total cohort at the start of the model shows that the calculation yields the correct number of patients over time.

c. Why in some cases (e.g. Markov_C!CK29) the utility value of a single cycle are halved instead of calculating the average of two subsequent cycles.

UCB response:

In developing the response to question B.15 a), a minor error was noted in the calculations of average state occupancy in the model (i.e. Markov_C!CK29). This error was corrected so that the average state occupancy was calculated as the average number of patients between the start and end of each cycle period. The only exception is in the calculation for average state occupancy in the first subsequent treatment state, Markov_C!CK28, which is based on an average of non-responders from Markov_C!K27, whom are assumed to instantly transition to the first subsequent treatment state at the end of the previous cycle, and the numbers occupying the state at the end of the cycle, Markov_C!W28. This is explained further in response to B15 a).

A more detailed explanation of the revisions to the executable model is provided in Appendix 3. The updates to the model were checked by an independent modeller, and extreme value tests were performed to ensure the model produced consistent and logical results.

The basecase results of the corrected model (Appendix 3) are consistent with the original base case submitted, such that the discrepancy in the above calculations had limited impact on the estimated ICERs, and had no impact on the conclusions of the base case analysis, which demonstrates that CZP is a cost-effective therapy in TNF-IR.

The revised deterministic and probabilistic results for the base case analysis are included in Appendix 3. It is expected that the results of the sensitivity analyses will be similar to those presented in the original UCB submission, given that there were only minor differences in base case results between the original and revised models (Appendix 3, Table 1).

B.16 Please clarify why the 80 mg dose of tocilizumab was not considered in the model.

UCB response:

The 200mg dose of tocilizumab was selected as the median dose available on the British National Formulary (ie. 80 mg, 200 mg, 400 mg).

B.17 Please clarify why in Figure 57 (page 172 of the company submission) the arrows from the two "Continue Tx HAQ change = d" states in the fourth row ("Cycle 3+;time=1.5 years") point at the "FU treatment 1* (2nd cycle+)" state instead of "FU treatment 1* (1st cycle)".

UCB response:

The model schematic Figure 57 from the UCB submission is intended to provide an illustration of transitions in the model. The row corresponding to time points 1.5 onwards represents all future transitions. For ease of presentation, the connecting nodes for response to FU treatment 1 (1st cycle) were omitted from the diagram. In the executable model, it is necessary for patients to first transition to the FU treatment 1 (1st cycle) state before entering the FU treatment 1 (2nd cycle +) state.

B.18 Please clarify the apparent discordance between the percentage of good or moderate responders reported by Emery et al. in the methotrexate treatment arm (16.5%) and the probability of inadequate response (83.7%) reported for the RADIATE trial in the company submission (page 200). It is believed the addition of the two values should equal 100%.

UCB response:

The 83.7% probability of inadequate response for subsequent therapies (including the MTX arm of RADIATE) reported in the UCB submission on page 200, were estimated from the NMA model, so as to provide a mean and associated standard error statistic that can be used in the probabilistic analysis. The probabilities of response generated in the NMA model are influenced by the weak/vague priors assigned to the trial-specific baseline effects, and hence may lead to the slight discrepancy between reported and modelled values (noting that the summation of the two values is 100.2%).

B.19 Please clarify why utility values were not age-adjusted.

UCB response:

Utility values considered in the submitted economic model were not subject to direct age-adjustment, as, to our knowledge, there are no data on the effect of age on utility in patients with established RA who have previous TNFi exposure. Furthermore, due to the modelled relationship between HAQ-DI and utility, the average utility of the cohort is simulated to decrease with time, in line with the predicted age-related worsening (ie. progression) in HAQ-DI. The model therefore indirectly accounts for the effects of aging on utility, as captured through the effects of age on HAQ-DI score.

B.20 Please clarify why biosimilars for etanercept were not considered.

UCB response:

The biosimilar to etanercept was launched in the UK on 16th February 2016. At the time of the UCB submission, there was no published information on the acquisition cost of this biosimilar. Hence, in the absence of data, biosimilars to etanercept were not considered in the submitted model.

B.21 Please provide a sensitivity analysis using the change in HAQ associated with good, moderate and no EULAR response as used in NICE technology appraisal 375. These values are 0.672, 0.317 and 0 respectively.

UCB response:

The results of the sensitivity analysis requested are presented in Table R (population A), Table S (population B), and Table T (population C).

Across all populations, the ICERs presented in this new sensitivity analysis are consistent with the results and conclusions of the base case analysis submitted by UCB, and hence show the robustness of the model to assumptions on the change in HAQ-DI associated with response.

Table R: Population A Sensitivity analysis results with HAQ change from TA375

Technologies (branded biosimilar)	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER (£/QALY) versus baseline	ICER (£/QALY) incremental
RTX + MTX (MabThera)	£138,186	16.135	6.998					Cost effective at WTP < £34,635
CZP+ MTX (Cimzia)	£148,105	16.240	7.284	£9,919	0.105	0.286	£34,635	**Cost effective at WTP > £34,635**

ICER, incremental cost effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life years

Table S: Population B Sensitivity analysis results with HAQ change from TA375

Technologies	Total costs (£)	Total LYG	Total QALYs	Incremental costs versus baseline (£)	Incremental LYG versus baseline	Incremental QALYs versus baseline	ICER (£/QALY) versus baseline	ICER (£/QALY) incremental
GOL + MTX (Simponi)	£95,976	15.889	6.045					**Cost effective at WTP < £4,637**
ADA + MTX (Humira)	£95,976	15.889	6.045	£0	0.000	0.000	-	
ETA + MTX (Enbrel)	£96,400	15.889	6.045	£424	0.000	0.000	Dominated by ADA/GOL	Dominated by ADA/GOL
CZP + MTX (Cimzia)	£97,181	15.925	6.305	£1,205	0.036	0.26	£4,637	**Cost effective at WTP between £4,637 and £137,138**
IFX + MTX (Remicade)	£100,277	15.889	6.045	£4,301	0.000	0.000	Dominated by ADA/GOL	Dominated by ADA/GOL

Technologies	Total costs (£)	Total LYG	Total QALYs	Incremental costs versus baseline (£)	Incremental LYG versus baseline	Incremental QALYs versus baseline	ICER (£/QALY) versus baseline	ICER (£/QALY) incremental
ABA + MTX (IV - no PAS) (Orencia)	£114,399	15.896	6.092	£18,423	0.007	0.047	£393,482	Dominated by CZP
TOC + MTX (IV - no PAS) (RoActemra)	£124,371	15.953	6.503	£28,395	0.064	0.458	£61,985	**Cost effective at WTP > £137,138**

ICER, incremental cost effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life years

Table T: Population C Sensitivity analysis results with HAQ change from TA375

Technologies	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER (£/QALY) versus baseline	ICER (£/QALY) incremental
ADA (Humira)	£94,542	15.845	5.874					**Cost effective at WTP < £7,127**
ETA (Enbrel)	£94,946	15.845	5.874	£404	0.000	0.000	Dominated by ADA	Dominated by ADA
CZP (Cimzia)	£96,390	15.882	6.133	£1,848	0.037	0.259	£7,127	**Cost effective at WTP between £7,127 and £129,102**
TOC (IV – no PAS) (RoActemra)	£122,859	15.910	6.338	£28,317	0.066	0.464	£60,977	**Cost effective at WTP > £129,102**

ICER, incremental cost effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life years

B.22 Please provide a sensitivity analysis in relation to the costs of palliative care.

UCB response:

A summary of the results of the requested sensitivity analysis are presented in Table U below. A detailed breakdown of results is provided in Appendix 2 of this response document.

Across all populations, the ICERs presented in the sensitivity analysis were largely consistent with the results and conclusions of the base case analysis submitted by UCB, and hence show the robustness of the model to assumptions on the costs of palliative care.

Table U: Results of sensitivity analysis assuming +/- 50% variation in resource use for palliative care

Parameter	Base case estimate	Sensitivity estimate	Population A results (ICER CZP vs. comparator)	Population B results (Fully incremental analysis)							Population C results (Fully incremental analysis)			
			Base case	CZP + MTX	ABA + MTX	ADA + MTX	ETA + MTX	GOL + MTX	IFX + MTX	TOC + MTX	CZP	ADA	ETA	TOC
			£34,516	CE at WTP of £3k to £129k	Dominated	CE at WTP < £3k	Dominated	CE at WTP < £3k	Dominated	CE at WTP > £129k	CE at WTP of £5k to £123k	CE at WTP < £5k	Dominated	CE at WTP > £123k
Palliative care	£900	+50% resource use	£32,254	CE at WTP of £2.5k to £128k	Dominated	CE at WTP < £2.5k	Dominated	CE at WTP < £2.5k	Dominated	CE at WTP > £128k	CE at WTP of £3.5k to £123k	CE at WTP < £3.5k	Dominated	CE at WTP > £123k
		-50% resource use	£36,778	CE at WTP of £5k to £130k	Dominated	CE at WTP < £5k	Dominated	CE at WTP < £5k	Dominated	CE at WTP > £130k	CE at WTP of £6k to £125k	CE at WTP < £6k	Dominated	CE at WTP > £125k

C. Textual clarifications and additional points

C.1 Please clarify whether the phrase waste within “unopened vials” being lost is a typographical error. From the model it would appear that leftover drugs in open vials are assumed to be lost, rather than unopened vials. Please clarify why the budget impact model assumes no vial wastage in contrast to the economic model.

UCB response:

Thank you for the comment. Yes, this is a typographical error; the text should read “opened vials”. The budget impact calculations were based on a simplified cost calculation approach, which assumed no vial wastage at a national-level. By assuming no vial wastage, the costs of IV therapies may be underestimated in the analysis, and hence the net cost-savings for certolizumab pegol may be underestimated in this population.

C.2 Table 13 (page 73 of the company submission): Please confirm that the percentages within the TNFi experienced patients sub-row in the REALISTIC study are incorrect as they have been divided by the wrong number of patients. There are several other instances within Table 13 of apparent miscalculation, please check the values throughout this table.

UCB response:

A revised version of Table 13 from UCB's submission is provided below, where all instances of miscalculation have been checked and updated, including the RF-positive data for REALISTIC, J-RAPID and HIKARI. Please refer to this version for all data instead of the previous version in the original submission.

"Table 13: Baseline characteristics of participants in the studies across treatment groups

		(n) [% TNFi experienced of total]	Mean age (SD), years	Female, n (%)	Mean disease duration, years (SD)	HAQ-DI mean (SD)	DAS28(ESR), mean (SD)	RF-positive (≥14 IU/mL), n (%)
REALISTIC[10] (NCT00717236)	All subjects	Overall patients (n=1,063)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		TNFi experienced (n=400) [37.6%]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	CZP 200 mg Q2W +/- MTX/cDMARDs	Overall patients (n=851)	55.4 (12.4)	660 (77.6)	8.6 (8.8)	1.5 (0.6)	6.4 (0.9)	555 (73.9)
		TNFi experienced (n=320) [37.6%]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	PBO +/- MTX/cDMARDs	Overall patients (n=212)	53.9 (12.7)	169 (79.7)	8.9 (9.1)	1.6 (0.6)	6.4 (0.9)	137 (76.5)
		TNFi experienced (n=80) [37.7%]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
DOSEFLEX[11] (NCT00580840)	All subjects‡	Overall patients (n=333)	54.2 (12.8)	76.0	6.4 (4.5)	1.52 (0.64)	6.4 (1.0)	315 (94.6)
		TNFi experienced (n=178) [53.5%]	54.2 (12.07)	77.0	7.6 (4.4)	1.6 (0.6)	6.4 (0.9)	167 (93.8)
	CZP 200 mg Q2W + MTX‡	Overall patients (n=70)	55.6 (10.7)	49 (70.0)	5.9 (4.2)	1.6 (0.7)	6.4 (0.8)	65 (92.9)
		TNFi experienced (n=43) [61.4%]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	CZP 400 mg Q4W + MTX‡	Overall patients (n=70)	53.1 (13.8)	58 (82.9)	6.4 (4.7)	1.4 (0.6)	6.2 (1.0)	65 (92.9)
		TNFi experienced (n=39) [55.7%]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	PBO + MTX‡	Overall patients (n=69)	51.5 (13.2)	56 (81.2)	6.5 (4.6)	1.4 (0.6)	6.4 (1.0))	69 (97.1)

		(n) [% TNFi experienced of total]	Mean age (SD), years	Female, n (%)	Mean disease duration, years (SD)	HAQ-DI mean (SD)	DAS28(ESR), mean (SD)	RF-positive (≥14 IU/mL), n (%)
		TNFi experienced (n=29) [42.0%]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
PREDICT[8] (NCT01255761)	CZP 200 mg Q2W +/- MTX/cDMARDs	Overall patients (n=733)	54.9	571 (77.9)	8.9 (9.1)	5.9†	6.3 (1.1)	493 (71.1)
		TNFi experienced (n=407) [55.5%]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		RAPID-3 assigned (n=368) [52.7%]	54.0	279 (75.8)	8.8 (9.3)	6.1†	6.3 (1.1)	251 (72.3)
		CDAI assigned (n=365) [58.4%]	55.7	292 (80.0)	9.1 (8.9)	5.8†	6.3 (1.1)	242 (69.9)
SWITCH[4] (NCT01147341)	CZP 200 mg Q2W + cDMARDs	TNFi experienced (n=27) [100.0%]	56.1	NR	12.0	1.5	5.5*	NR
	PBO + cDMARDs	TNFi experienced (n=10) [100.0%]	59.0	NR	14.0	1.1	5.4*	NR
J-RAPID[12] (NCT00791999)	CZP 200 mg Q2W + MTX	Overall patients (n=82[redacted]	50.6 (11.4)	69 (84.1)	5.6 (4.2)	1.1 (0.7)	6.2 (0.8)	71 (86.6)
	PBO + MTX	Overall patients (n=77) [redacted]	51.9 (11.1)	66 (85.7)	5.8 (4.1)	1.2 (0.7)	6.5 (0.9)	66 (85.7)
HIKARI[13] (NCT00791921)	CZP 200 mg Q2W -/+ non-MTX cDMARDs≠	Overall patients (n=116) [6.9%]	56.0 (10.2)	83.7 (71.6)	5.4 (4.0)	1.05 (0.7)	6.1 (0.9)	99 (85.3)
	PBO -/+ non-MTX cDMARDs	Overall patients (n=114) [14.0%]	55.4 (9.8)	88 (77.2)	5.8 (4.3)	1.21 (0.7)	6.3 (1.0)	102 (89.5)

SD: standard deviation; NR: not reported; CV: coefficient of variation; RAPID-3: routine assessment of patient index data 3; CDAI: clinical disease activity index. ∓For REALISTIC, selected baseline characteristics were only recorded in a subset of the patients within the overall study population, and are indicated as [n] where appropriate. ‡For DOSEFLEX, baseline characteristics for "all subjects" refers to all subjects in the modified enrolled set who entered the 4 week run-in phase, while the PBO and CZP stratification data represent patients who completed the run-in phase and were subsequently randomised into the three treatment groups (PBO, 200 mg CZP or 400 mg CZP) for the double-blind phase. †For PREDICT, MD-HAQ Global scores at baseline (within a range of 0-10) are presented. *For SWITCH, DAS28(CRP) at baseline is presented, DAS28(ESR) was not measured. ≠CZP in combination with non-MTX cDMARDs is not approved in the European Union."

C.3 Table 24 (page 109 of the company submission): The percentage of patients in remission in the certolizumab pegol treatment arm appears to be incorrectly calculated, please confirm if this is correct.

UCB response:

UCB would kindly ask the ERG to clarify their question. Table 24 (page 109 of UCB's submission) does not report percentages of patients in remission. Table 23 (page 106 of UCB's submission) does report such data, however, the data in this table appear to be correct.

C.4 Please update Figure 2 (page 41 of the company submission) so that NICE technology appraisal 375 is taken into consideration. This would include renaming some of the technology appraisals and adding tocilizumab monotherapy as a first-line bDMARD option. This latter point may mean that the heading of one box may need to be changed.

UCB response:

The image in Figure 2 of UCB's submission is the RA treatment algorithm downloaded from the NICE website reflecting the RA pathway as discussed during the final scope of the STA. If any of the requested updates were to be implemented, these would reflect UCB's own interpretation of the changes to the pathway and therefore would not represent NICE's update of the treatment process as per TA375. UCB would suggest retaining the current Figure as submitted, unless an updated pathway can be provided by NICE.

C.5 Rendas-Baum et al. report that more than 90% of rheumatologists switched patients to alternative TNFi therapy yet page 38 suggests 94%. Please clarify from where the more precise value was identified.

UCB response:

Thank you for pointing out the difference. The value of 90% as per the original reference (Rendas-Baum et al.) should be considered correct, thus please disregard the value of 94% from page 38 of UCB's submission.

C.6 Please clarify whether the mean length of the first anti-TNFα therapy in Hyrich et al. was 13 months rather than the 6 months reported in page 38 of the company submission.

UCB response:

Thank you for pointing out the difference. The value of 13 months as per the original reference (Hyrich et al.) should be considered correct, thus please disregard the value of 6 months from page 38 of UCB's submission.

C.7 Please clarify whether 16% of patients who discontinued first biologic agent due to inefficacy also discontinued second biologic agent due to inefficacy, as reported in Hyrich et al, instead of 13% as reported on page 38 of the company submission.

UCB response:

Thank you for pointing out the difference. The value of 16% as per the original reference (Hyrich et al.) should be considered correct, thus please disregard the value of 13% from page 38 of UCB's submission.

References

1. Simard JF, Arkema EV, Sundstrom A, et al. Ten years with biologics: to whom do data on effectiveness and safety apply? Rheumatology (Oxford). 2011;50(1):204-13.
2. Curtis JR,de Longueville M,O'Brien C and Haraoui B. Improvement in Disease Activity and the Long-Term Risk of Serious Infectious Events in Rheumatoid Arthritis Patients Treated with Certolizumab Pegol [Abstract 561]. Arthritis Rheumatol. 2015;67(Suppl 10).
3. Yun H, Xie F, Delzell E, et al. Comparative Risk of Hospitalized Infection Associated With Biologic Agents in Rheumatoid Arthritis Patients Enrolled in Medicare. Arthritis Rheumatol. 2016;68(1):56-66.
4. Schiff MH, von Kempis J, Goldblum R, et al. Rheumatoid arthritis secondary non-responders to TNF can attain an efficacious and safe response by switching to certolizumab pegol: a phase IV, randomised, multicentre, double-blind, 12-week study, followed by a 12-week open-label phase. Ann Rheum Dis. 2014;73(12):2174-7.
5. Bykerk VP, Cush J, Winthrop K, et al. Update on the safety profile of certolizumab pegol in rheumatoid arthritis: an integrated analysis from clinical trials. Ann Rheum Dis. 2015;74(1):96-103.
6. Yamamoto K, Takeuchi T, Yamanaka H, et al. Efficacy and safety of certolizumab pegol plus methotrexate in Japanese rheumatoid arthritis patients with an inadequate response to methotrexate: the J-RAPID randomized, placebo-controlled trial. Modern Rheumatology. 2014;24(5):715-724.
7. Yamamoto K, Takeuchi T, Yamanaka H, et al. Efficacy and safety of certolizumab pegol without methotrexate co-administration in Japanese patients with active rheumatoid arthritis: the HIKARI randomized, placebo-controlled trial. Modern Rheumatology. 2014;24(4):552-560.
8. Curtis JR, Churchill M, Kivitz A, et al. A Randomized Trial Comparing Disease Activity Measures to Assess and Predict Response in Rheumatoid Arthritis Patients Initiating Certolizumab Pegol. Arthritis Rheumatol. 2015.
9. Dias S,Welton NJ,Sutton AJ and Ades A. NICE DSU technical support document 2: a generalised linear modelling framework for pairwise and network meta-analysis of randomised controlled trials. URL http://www.nicedsu.org.uk. 2011.
10. Weinblatt ME, Fleischmann R, Huizinga TW, et al. Efficacy and safety of certolizumab pegol in a broad population of patients with active rheumatoid arthritis: results from the REALISTIC phase IIIb study. Rheumatology (Oxford). 2012.
11. Furst DE, Shaikh SA, Greenwald M, et al. Two dosing regimens of certolizumab pegol in patients with active rheumatoid arthritis. Arthritis Care Res (Hoboken). 2015;67(2):151-60.
12. Yamanaka H, Yamamoto K, Takeuchi T, et al. Improved Physical Function, Pain, and Health Related Quality of Life with Certolizumab Pegol in Japanese Rheumatoid Arthritis Patients with an Inadequate Response to Methotrexate: Results from the JRAPID Study. Ann Rheum Dis. 2012;71(Suppl3):664.
13. Yamanaka H, Yamamoto K, Takeuchi T, et al. Improved Physical Function, Pain, and Health Related Quality of Life With Certolizumab Pegol in Japanese Rheumatoid Arthritis Patients without Methotrexate Co-Administration: Results from the HIKARI Study. Ann Rheum Dis. 2012;71(Suppl3):664.
14. Brennan A, Bansback N, Nixon R, et al. Modelling the cost effectiveness of TNF-alpha antagonists in the management of rheumatoid arthritis: results from the British Society for Rheumatology Biologics Registry. Rheumatology (Oxford). 2007;46(8):1345-54.
15. Bansback N, Marra C, Tsuchiya A, Anis A, Guh D, Hammond T, Brazier J. Using the Health Assessment Questionnaire to Estimate Preference-Based Single Indices in Patients With Rheumatoid Arthritis. Arthritis & Rheumatism 57(6): 963-971; 2007

Appendices

1 Appendix 1: Table for Response A.3

Table V: Summary of mapping publications listed in the HERC mapping database

Citation details	Year of publication	Quality of life measures		Disease or patient group	Disease category	No. observations in estimation sample	Related papers and resources	Included in Pennington & Davis?
		From	To					
Adams R., Walsh C., Veale D., Bresnihan B., FitzGerald O., Barry M. (2010). Understanding the relationship between the EQ-5D, SF-6D, HAQ and disease activity in inflammatory arthritis. Pharmacoeconomics. 28 (6), 477-87.	2010	HAQ	EQ-5D	RA	Musculo-skeletal	345 pts	Models re-estimated in Adams, R et al. (2011). Value Health. 14, 921-7.	Yes
		HAQ	SF-6D	RA	Musculo-skeletal	345 pts	Models re-estimated in Adams, R et al. (2011). Value Health. 14, 921-7.	Yes
Adams R., Craig B. M., Walsh C. D., Veale D. J., Bresnihan B., FitzGerald O., et al. (2011). The impact of a revised EQ-5D population scoring on preference-based utility scores in an inflammatory arthritis cohort. Value Health. 14 (6), 921-7.	2011	HAQ	EQ-5D	RA	Musculo-skeletal	345 pts	Models predicting the standard EQ-5D tariff were previously reported in Adams, R et al. (2010). Pharmacoeconomics. 28, 477-87.	No – however, study linked to Adams 2010
Bansback N, Marra C, Tsuchiya A, Anis A, Guh D, Hammond T, et al. Using the health assessment questionnaire to estimate preference-based single indices in patients with rheumatoid arthritis. Arthritis Rheum. 2007 Aug 15;57(6):963-71.	2007	HAQ-DI	EQ-5D	RA	Musculo-skeletal	439	Versteegh (Versteegh, MM et al. (2010). Health Qual Life Outcomes. 8, 141) externally validated and used to explore how poor predictions are for those in poor health.	Yes
Barton, P., Jobanputra, P., Wilson, J., Bryan, S., & Burls, A. (2004). The use of modelling to evaluate new drugs for patients with a chronic condition: the case of antibodies against tumour necrosis factor in rheumatoid arthritis. Health Technol Assess, 8(11), iii, 1-91.	2004	HAQ	EQ-5D	RA	Musculo-skeletal	233	Mapping described on page 22-3. Uses the same dataset as Hurst 1997 (Br J Rheumatol, 36, 551-559). Pennington et al (2014, Value Health, 17, 762-771)	Yes

Citation details	Year of publication	Quality of life measures		Disease or patient group	Disease category	No. observations in estimation sample	Related papers and resources	Included in Pennington & Davis?
		From	To					
							compare performance against different mapping algorithms and assess impact on an economic model.	
Carreno A., Fernandez I., Badia X., Varela C., Roset M. (2011). Using HAQ-DI to estimate HUI-3 and EQ-5D utility values for patients with rheumatoid arthritis in Spain. Value Health. 14 (1), 192-200.	2011	HAQ-DI	EQ-5D	RA	Musculo-skeletal	235		Yes
Hawthorne, G., Buchbinder, R., & Defina, J. (2000). Functional status and health-related quality of life assessment in patients with rheumatoid arthritis. Centre for Health Program Evaluation, working paper 116.	2000	HAQ	EQ-5D	RA	Musculo-skeletal	139	Pennington et al (2014, Value Health, 17, 762-771) compare performance against different mapping algorithms and assess impact on an economic model.	Yes
		HAQ	AQoL	RA	Musculo-skeletal	139		
Hernández Alava M., Wailoo A. J., Ara R. (2012). Tails from the peak district: adjusted limited dependent variable mixture models of EQ-5D questionnaire health state utility values. Value Health. 15 (3), 550-61.	2012	HAQ-DI	EQ-5D	RA	Musculo-skeletal	467 patients	Pre-publication version including appendix giving methods for calculating predictions available at: http://www.sheffield.ac.uk/polopoly_fs/1.215354!/file/10.08.pdf. Covariance matrix available at http://www.sheffield.ac.uk/scharr/sections/heds/dps-2010.	Yes
Hernández Alava, M., Wailoo, A., Wolfe, F., & Michaud, K. (2013). The relationship between EQ-5D, HAQ and pain in patients with rheumatoid arthritis. Rheumatology (Oxford), 52, 944-950; Hernandez Alava, M., Wailoo, A., Wolfe, F., & Michaud, K. (2013). A Comparison of Direct and Indirect Methods for the Estimation of Health Utilities from Clinical Outcomes. Med Decis Making; and Hernández Alava, M., Wailoo, A., Wolfe, F., & Michaud, K. (2012). A comparison of direct and indirect methods for the estimation of health utilities from clinical outcomes. University of Sheffield, HEDS Discussion Paper Retrieved 3rd December 2012, from http://www.nicedsu.org.uk/Mapping%20of%20EQ-5D.DP.pdf	2012-2013	HAQ and pain on VAS	EQ-5D	RA	Musculo-skeletal	100,398	Covariance matrix available at http://rheumatology.oxfordjournals.org/content/52/5/944/suppl/DC1. Discussion paper also gives coefficients for response mapping model.	Yes
Hurst, N. P., Kind, P., Ruta, D., Hunter, M., & Stubbings, A. (1997). Measuring health-related quality of life in rheumatoid arthritis: validity, responsiveness and reliability of EuroQol (EQ-5D). Br J Rheumatol, 36(5), 551-559.	1997	HAQ, pain visual acuity, ACR disease activity	EQ-5D	RA	Musculo-skeletal	233		Yes

Citation details	Year of publication	Quality of life measures		Disease or patient group	Disease category	No. observations in estimation sample	Related papers and resources	Included in Pennington & Davis?
		From	To					
		and clinical measures						
Kobelt, G., Jonsson, L., Lindgren, P., Young, A., & Eberhardt, K. (2002). Modeling the progression of rheumatoid arthritis: a two-country model to estimate costs and consequences of rheumatoid arthritis. Arthritis Rheum, 46(9), 2310-2319. Kobelt, G., Lindgren, P., Lindroth, Y., Jacobson, L., & Eberhardt, K. (2005). Modelling the effect of function and disease activity on costs and quality of life in rheumatoid arthritis. Rheumatology (Oxford), 44(9), 1169-1175. Kobelt, G., Lindgren, P., Singh, A., & Klareskog, L. (2005). Cost effectiveness of etanercept (Enbrel) in combination with methotrexate in the treatment of active rheumatoid arthritis based on the TEMPO trial. Ann Rheum Dis, 64(8), 1174-1179.	2002/2005	HAQ	EQ-5D	RA	Musculo-skeletal	519	Pennington et al (2014, Value Health, 17, 762-771) compare performance against different mapping algorithms and assess impact on an economic model.	Yes
Lindgren, P., Geborek, P., & Kobelt, G. (2009). Modeling the cost-effectiveness of treatment of rheumatoid arthritis with rituximab using registry data from Southern Sweden. Int J Technol Assess Health Care, 25(2), 181-189.	2009	HAQ and DAS28	EQ-5D	RA	Musculo-skeletal	6,860		Yes
Malottki, K., Barton, P., Tsourapas, A., Uthman, A. O., Liu, Z., Routh, K., Connock, M., Jobanputra, P., Moore, D., Fry-Smith, A., & Chen, Y. F. (2011). Adalimumab, etanercept, infliximab, rituximab and abatacept for the treatment of rheumatoid arthritis after the failure of a tumour necrosis factor inhibitor: a systematic review and economic evaluation. Health Technol Assess, 15(14), 1-278.	2011	HAQ	EQ-5D	RA	Musculo-skeletal	233	See pages 156 and 272 for details. Re-analyses the same dataset as Hurst, 1997 Br J Rheumatol, 36(5), 551-559.	Yes
Marra, C. A., Marion, S. A., Guh, D. P., Najafzadeh, M., Wolfe, F., Esdaile, J. M., Clarke, A. E., Gignac, M. A., & Anis, A. H. (2007). Not all "quality-adjusted life years" are equal. J Clin Epidemiol, 60(6), 616-624.	2007	HAQ	EQ-5D	RA	Musculo-skeletal	317	Pennington et al (2014, Value Health, 17(8), 762-771) compare performance against different mapping algorithms and assess impact on an economic model.	Yes
Michaud, K., & Wolfe, F. (2005). EQ5D changes rheumatoid arthritis quality of life in United States: A retrospective study of 11,289 patients. Arthritis Rheum, 52(Suppl), S400.	2005	HAQ	EQ-5D	RA	Musculo-skeletal	35,422	Regression coefficients given in Beresniak et al. (2007) J Rheumatol. 34, 2193-2200	No
Ota, H., Tanno, M., Tanaka, H., Kobayashi, M., & Yoshino, S. (2003). Correlation between the health assessment questionnaire (HAQ) and utility value in rheumatoid arthritis patients. Presented at the ISPOR first Asia-Pacific	2003/2006	HAQ	EQ-5D	RA	Musculo-skeletal	307	Pennington et al (2014, Value Health, 17, 762-771) compare performance against different mapping algorithms and assess	No

Citation details	Year of publication	Quality of life measures		Disease or patient group	Disease category	No. observations in estimation sample	Related papers and resources	Included in Pennington & Davis?
		From	To					
conference September 1-3; Kobe, Japan. Tanno, M., Nakamura, I., Ito, K., Tanaka, H., Ohta, H., Kobayashi, M., Tachihara, A., Nagashima, M., Yoshino, S., & Nakajima, A. (2006). Modeling and cost-effectiveness analysis of etanercept in adults with rheumatoid arthritis in Japan: a preliminary analysis. Mod Rheumatol, 16(2), 77-84.							impact on an economic model.	
Soini E. J., Hallinen T. A., Puolakka K., Vihervaara V., Kauppi M. J. (2012). Cost-effectiveness of adalimumab, etanercept, and tocilizumab as first-line treatments for moderate-to-severe rheumatoid arthritis. J Med Econ. 15 (2), 340-51 and Ducournau, P., Kielhorn, A., & Wintfeld, N. (2009). Comparison of linear and nonlinear utility mapping between HAQ and EQ-5D using pooled data from the tocilizumab trials OPTION and LITHE. Rheumatology (Oxford), 48(1 Suppl), i107-108.	2012	HAQ	EQ-5D	RA	Musculo-skeletal	1,812		No
Standfield, L., Norris, S., Harvey, C., Elliot, L., Riordan, J., Hall, S., Day, R., Nash, P., Thirunavukkarasu, K., Robertson, J., & Palmer, T. (2010). Relationship between rheumatoid arthritis disease severity, health-related utility, and resource use in Australian patients: A cross-sectional, multicenter study. Clin Ther, 32(7), 1329-1342.	2010	HAQ	EQ-5D	RA	Musculo-skeletal	169	Coefficients shown in Figure 1. Pennington et al (2014, Value Health, 17, 762-771) compare performance against different mapping algorithms and assess impact on an economic model.	Yes
		HAQ	HUI3	RA	Musculo-skeletal	170		
Vera-Llonch, M., Massarotti, E., Wolfe, F., Shadick, N., Westhovens, R., Sofrygin, O., Maclean, R., Li, T., & Oster, G. (2008). Cost-effectiveness of abatacept in patients with moderately to severely active rheumatoid arthritis and inadequate response to tumor necrosis factor-alpha antagonists. J Rheumatol, 35(9), 1745-1753. Vera-Llonch, M., Massarotti, E., Wolfe, F., Shadick, N., Westhovens, R., Sofrygin, O., Maclean, R., Yuan, Y., & Oster, G. (2008). Cost-effectiveness of abatacept in patients with moderately to severely active rheumatoid arthritis and inadequate response to methotrexate. Rheumatology (Oxford), 47(4), 535-541.	2008	HAQ	EQ-5D	RA	Musculo-skeletal	~19,000	Pennington et al (2014, Value Health, 17, 762-771) compare performance against different mapping algorithms and assess impact on an economic model.	Yes
Versteegh M. M., Leunis A., Luime J. J., Boggild M., Uyl-de Groot C. A., Stolk E. A. (2012). Mapping QLQ-C30, HAQ, and MSIS-29 on EQ-5D. Med Decis Making. 32 (4), 554-68.	2012	HAQ	EQ-5D	RA	Musculo-skeletal	186		No
		HAQ, SF-36, Hospital Anxiety and	EQ-5D	RA	Musculo-skeletal	186		

Citation details	Year of publication	Quality of life measures		Disease or patient group	Disease category	No. observations in estimation sample	Related papers and resources	Included in Pennington & Davis?
		From	To					
		Depression Scale (HADS) & DAS28						
Versteegh M. M., Rowen D., Brazier J. E., Stolk E. A. (2010). Mapping onto Eq-5 D for patients in poor health. Health Qual Life Outcomes. 8, 141.	2010	HAQ	EQ-5D	Patients with and without RA	Various	493	The authors themselves do not recommend using this algorithm to predict utilities as some participants had RA while others did not. Evaluate various other mapping models.	No
Wolfe F., Michaud K., Wallenstein G. (2010). Scale characteristics and mapping accuracy of the US EQ-5D, UK EQ-5D, and SF-6D in patients with rheumatoid arthritis. J Rheumatol. 37 (8), 1615-25.	2010	HAQ	EQ-5D	RA	Musculo-skeletal	10,895		No

2 Appendix 2: Tables for Response B.22

Table W: Population A with palliative care resource increased by 50%

Technologies (branded biosimilar)	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER (£/QALY) versus baseline	ICER (£/QALY) incremental
RTX + MTX (MabThera)	£143,616	16.132	6.975					Cost effective at WTP < £32,254
CZP+ MTX (Cimzia)	£152,902	16.237	7.263	£9,287	0.105	0.288	£32,254	**Cost effective at WTP > £32,254**

ICER, incremental cost effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life years

Table X: Population A with palliative care resource decreased by 50%

Technologies (branded biosimilar)	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER (£/QALY) versus baseline	ICER (£/QALY) incremental
RTX + MTX (MabThera)	£135,433	16.132	6.975					Cost effective at WTP < £36,778
CZP+ MTX (Cimzia)	£146,022	16.237	7.263	£10,589	0.105	0.288	£36,778	**Cost effective at WTP > £36,778**

ICER, incremental cost effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life years

Table Y: Population B with palliative care resource increased by 50%

Technologies	Total costs (£)	Total LYG	Total QALYs	Incremental costs versus baseline (£)	Incremental LYG versus baseline	Incremental QALYs versus baseline	ICER (£/QALY) versus baseline	ICER (£/QALY) incremental
GOL + MTX (Simponi)	£104,850	15.881	6.016					**Cost effective at WTP < £2,537**
ADA + MTX (Humira)	£104,850	15.881	6.016	£0	0.000	0.000	-	
ETA + MTX (Enbrel)	£105,274	15.881	6.016	£423	0.000	0.000	Dominated by ADA/GOL	Dominated by ADA/GOL
CZP + MTX (Cimzia)	£105,534	15.922	6.286	£684	0.041	0.270	£2,537	**Cost effective at WTP between £2,537 and £128,355**
IFX + MTX (Remicade)	£109,150	15.881	6.016	£4,300	0.000	0.000	Dominated by ADA/GOL	Dominated by ADA/GOL
ABA + MTX (IV - no PAS) (Orencia)	£122,810	15.889	6.065	£17,960	0.007	0.049	£369,765	Dominated by CZP
TOC + MTX (IV - no PAS) (RoActemra)	£131,858	15.953	6.491	£27,008	0.072	0.475	£56,903	**Cost effective at WTP > £128,355**

ICER, incremental cost effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life years

Table Z: Population B with palliative care resource decreased by 50%

Technologies	Total costs (£)	Total LYG	Total QALYs	Incremental costs versus baseline (£)	Incremental LYG versus baseline	Incremental QALYs versus baseline	ICER (£/QALY) versus baseline	ICER (£/QALY) incremental
GOL + MTX (Simponi)	£90,336	15.881	6.016					**Cost effective at WTP < £4,744**
ADA + MTX (Humira)	£90,336	15.881	6.016	£0	0.000	0.000	-	
ETA + MTX (Enbrel)	£90,760	15.881	6.016	£423	0.000	0.000	Dominated by ADA/GOL	Dominated by ADA/GOL
CZP + MTX (Cimzia)	£91,615	15.922	6.286	£1,279	0.041	0.270	£4,744	**Cost effective at WTP between £4,744 and £130,282**
IFX + MTX (Remicade)	£94,637	15.881	6.016	£4,300	0.000	0.000	Dominated by ADA/GOL	Dominated by ADA/GOL
ABA + MTX (IV - no PAS) (Orencia)	£108,409	15.889	6.065	£18,072	0.007	0.049	£372,076	Dominated by CZP
TOC + MTX (IV - no PAS) (RoActemra)	£118,334	15.953	6.491	£27,998	0.072	0.475	£58,989	**Cost effective at WTP > £130,282**

ICER, incremental cost effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life years

Table AA: Population C with palliative care resource increased by 50%

Technologies	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER (£/QALY) versus baseline	ICER (£/QALY) incremental
ADA (Humira)	£104,339	15.844	5.845					**Cost effective at WTP < £3,683**
ETA (Enbrel)	£104,744	15.844	5.845	£404	0.000	0.000	Dominated by ADA	Dominated by ADA
CZP (Cimzia)	£105,336	15.887	6.115	£997	0.043	0.271	£3,683	**Cost effective at WTP between £3,683 and £122,771**
TOC (IV – no PAS) (RoActemra)	£131,495	15.920	6.328	£27,156	0.076	0.484	£56,142	**Cost effective at WTP > £122,771**

ICER, incremental cost effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life years

Table BB: Population C with palliative care resource decreased by 50%

Technologies	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER (£/QALY) versus baseline	ICER (£/QALY) incremental
ADA (Humira)	£87,546	15.844	5.845					**Cost effective at WTP < £6,287**
ETA (Enbrel)	£87,951	15.844	5.845	£404	0.000	0.000	Dominated by ADA	Dominated by ADA
CZP (Cimzia)	£89,248	15.887	6.115	£1,701	0.043	0.271	£6,287	**Cost effective at WTP between £6,287 and £125,058**
TOC (IV – no PAS) (RoActemra)	£115,894	15.920	6.328	£28,348	0.076	0.484	£58,607	**Cost effective at WTP > £125,058**

ICER, incremental cost effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life years

3 Appendix 3: Responses to B.13 and B.15 (c) .

Two minor errors that were identified when developing responses to questions B.13 and B.15 c) from the ERG:

A more detailed explanation of the revisions to the Excel model (addressing the 2 above minor errors) is provided at the end of Appendix 3. The updates to the model were checked by an independent modeller, and extreme value tests were performed to ensure the model produced consistent and logical results.

The results of the revised base case are presented below for populations A, B and C. A comparison of the originally submitted versus revised base case results are summarised in Table V below.

The results of the revised base case are consistent with the original base case, such that the minor errors in cycle length and lifetable calculations had limited impact on the estimated ICERs, and no impact on the conclusions of the base case analysis, which demonstrates that CZP is a cost-effective therapy in TNF-IR.

Table CC: Basecase results presented in original UCB submission versus revised model

Population	Original submitted model	Revised model
Population A	Deterministic ICER: CZP versus rituximab: **£34,516** per QALY gained Probabilistic ICER: CZP versus rituximab: **£33,665** per QALY gained Probability that CZP is cost-effective at a threshold of £30,000 per QALY: **37.4%**	Deterministic ICER: CZP versus rituximab: **£34,378** per QALY gained Probabilistic ICER: CZP versus rituximab: **£33,222** per QALY gained Probability that CZP is cost-effective at a threshold of £30,000 per QALY:**37.0%**
Population B	Deterministic fully incremental ICER: GOL/ADA least costly ICER: CZP vs. GOL/ADA = **£3,641** ICER: TOC vs. CZP = **£129,316** ETN, IFX, ABA dominated or extendedly dominated Probability that CZP is cost-effective at a threshold of £30,000 per QALY: **95.9%**	Deterministic fully incremental ICER: GOL/ADA least costly ICER: CZP vs. GOL/ADA = **£3,527** ICER: TOC vs. CZP = **£135,953** ETN, IFX, ABA dominated or extendedly dominated Probability that CZP is cost-effective at a threshold of £30,000 per QALY: **96.0%**
Population C	Deterministic fully incremental ICER: ADA least costly ICER: CZP vs. ADA = **£4,985** ICER: TOC vs. CZP = **£123,915** ETN dominated Probability that CZP is cost-effective at a threshold of £30,000 per QALY: **97.2%**	Deterministic fully incremental ICER: ADA least costly ICER: CZP vs. ADA = **£6,213** ICER: TOC vs. CZP = **£127,955** ETN dominated Probability that CZP is cost-effective at a threshold of £30,000 per QALY: **97.5%**

The revised deterministic and probabilistic results for the base case analysis are provided below. We expect that the results of the sensitivity analyses will be similar to those presented in the original submission, given that there were only minor differences in base case results between the original and revised models (Table V).

3.1 Revised deterministic results

Table DD: Revised base-case cost effectiveness results in population A

Technologies (branded biosimilar)	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER (£/QALY) versus baseline	ICER (£/QALY) incremental
RTX + MTX (MabThera)	£138,520	16.139	7.000					Cost effective at WTP < £34,378
CZP+ MTX (Cimzia)	£148,361	16.244	7.286	£9,842	0.105	0.286	£34,378	**Cost effective at WTP > £34,378**

ICER, incremental cost effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life years

Table EE: Revised base-case cost effectiveness results in population B

Technologies	Total costs (£)	Total LYG	Total QALYs	Incremental costs versus baseline (£)	Incremental LYG versus baseline	Incremental QALYs versus baseline	ICER (£/QALY) versus baseline	ICER (£/QALY) incremental
GOL + MTX (Simponi)	£97,183	15.892	6.048					**Cost effective at WTP < £3,527**
ADA + MTX (Humira)	£97,183	15.892	6.048	£0	0.000	0.000	-	
ETA + MTX (Enbrel)	£97,606	15.892	6.048	£424	0.000	0.000	Dominated by ADA/GOL	Dominated by ADA/GOL
CZP + MTX (Cimzia)	£98,100	15.929	6.308	£918	0.038	0.260	£3,527	**Cost effective at WTP between £3,527 and £135,953**
IFX + MTX (Remicade)	£101,484	15.892	6.048	£4,301	0.000	0.000	Dominated by ADA/GOL	Dominated by ADA/GOL
ABA + MTX (IV - no PAS) (Orencia)	£115,555	15.898	6.095	£18,373	0.007	0.047	£392,027	Dominated by CZP
TOC + MTX (IV - no PAS) (RoActemra)	£125,112	15.958	6.507	£27,929	0.067	0.459	£60,869	**Cost effective at WTP > £135,953**

ICER, incremental cost effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life years

Table FF: Revised base-case cost effectiveness results in population C

Technologies	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER (£/QALY) versus baseline	ICER (£/QALY) incremental
ADA (Humira)	£95,632	15.856	5.880					**Cost effective at WTP < £6,213**
ETA (Enbrel)	£96,036	15.856	5.880	£404	0.000	0.000	Dominated by ADA	Dominated by ADA
CZP (Cimzia)	£97,249	15.895	6.141	£1,617	0.040	0.260	£6,213	**Cost effective at WTP between £6,213 and £127,955**
TOC (IV – no PAS) (RoActemra)	£123,592	15.926	6.346	£27,960	0.070	0.466	£59,973	**Cost effective at WTP > £127,955**

ICER, incremental cost effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life years

3.2 Revised probabilistic results

Results presented below are based on 5,000 iterations.

Table GG: Revised probabilistic sensitivity analysis results in population A

Therapy	Mean costs	Difference in costs (CZP vs. treatment)	Mean QALYs	Difference in QALYs (CZP vs. treatment)	ICER (CZP vs. treatment)	Probability of cost effectiveness at WTP threshold of £20,000/QALY (%)	Probability of cost effectiveness at WTP threshold of £30,000/QALY (%)
CZP + MTX	£149,579	£9,647	7.321	0.290	£33,222	2.20%	36.98%
RTX + MTX	£139,933		7.031			97.80%	63.02%

Results represent mean estimates per patient

Figure D: Revised Cost effectiveness acceptability curves in population A

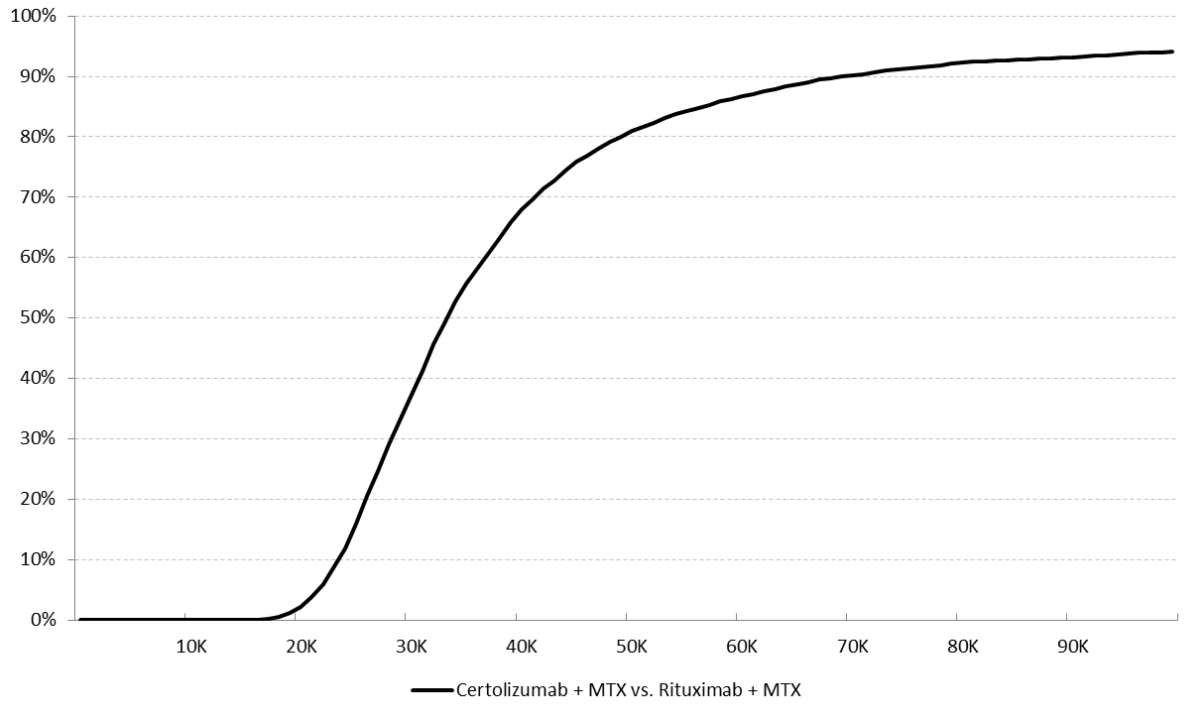


Table HH: Revised probabilistic sensitivity analysis results in population B

Therapy	Mean costs	Difference in costs (CZP vs. treatment)	Mean QALYs	Difference in QALYs (CZP vs. treatment)	ICER (CZP vs. treatment)	Probability of cost effectiveness at WTP threshold of £20,000/QALY (%)	Probability of cost effectiveness at WTP threshold of £30,000/QALY (%)
CZP + MTX	£98,848		6.327			99.5%	96.0%
ABA + MTX	£116,232	-£17,384	6.119	0.208	CZP dominates	0.0%	0.0%
ADA + MTX	£98,015	£833	6.076	0.251	£3,317	0.2%	1.7%
ETA + MTX	£98,360	£488	6.070	0.257	£1,900	0.0%	0.7%
GOL + MTX	£97,964	£885	6.071	0.256	£3,461	0.3%	1.5%
IFX + MTX	£102,242	-£3,394	6.070	0.257	CZP dominates	0.0%	0.0%
TOC + MTX	£125,507	-£26,658	6.528	-0.201	£132,783	0.0%	0.0%

**Original brands only (ie Remicade (IFX) and Enbrel (ETA))
CZP dominates comparator (ie. CZP is more effective and less costly)
Results represent mean estimates per patient

Figure E: Revised Cost effectiveness acceptability curves in population B

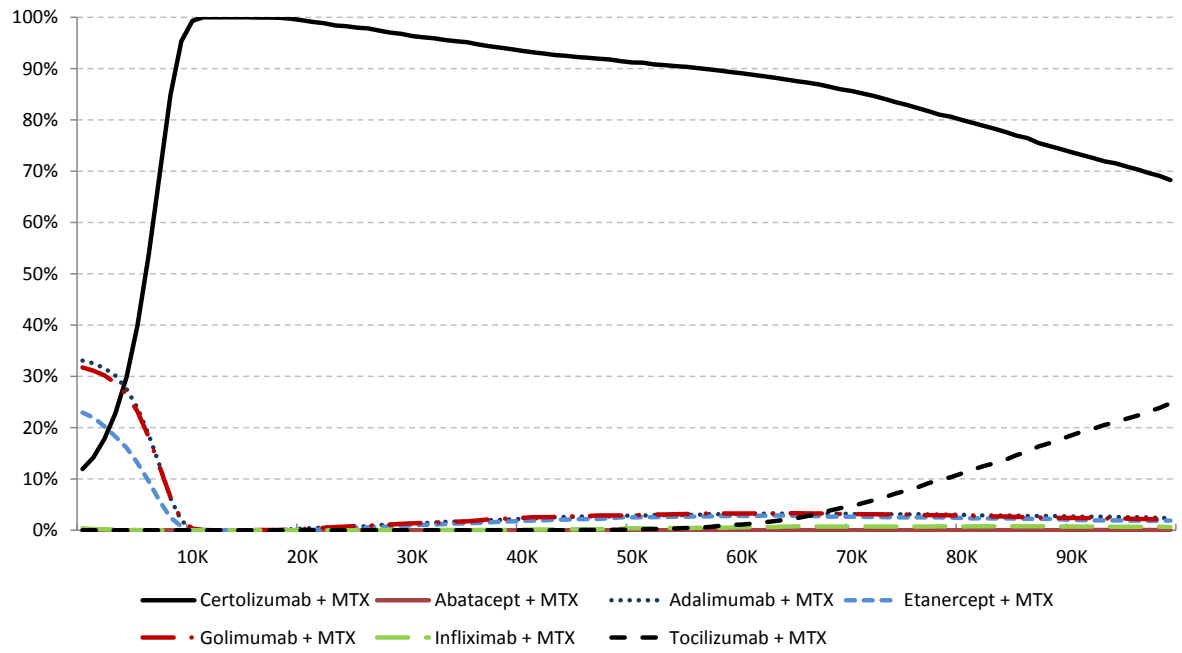
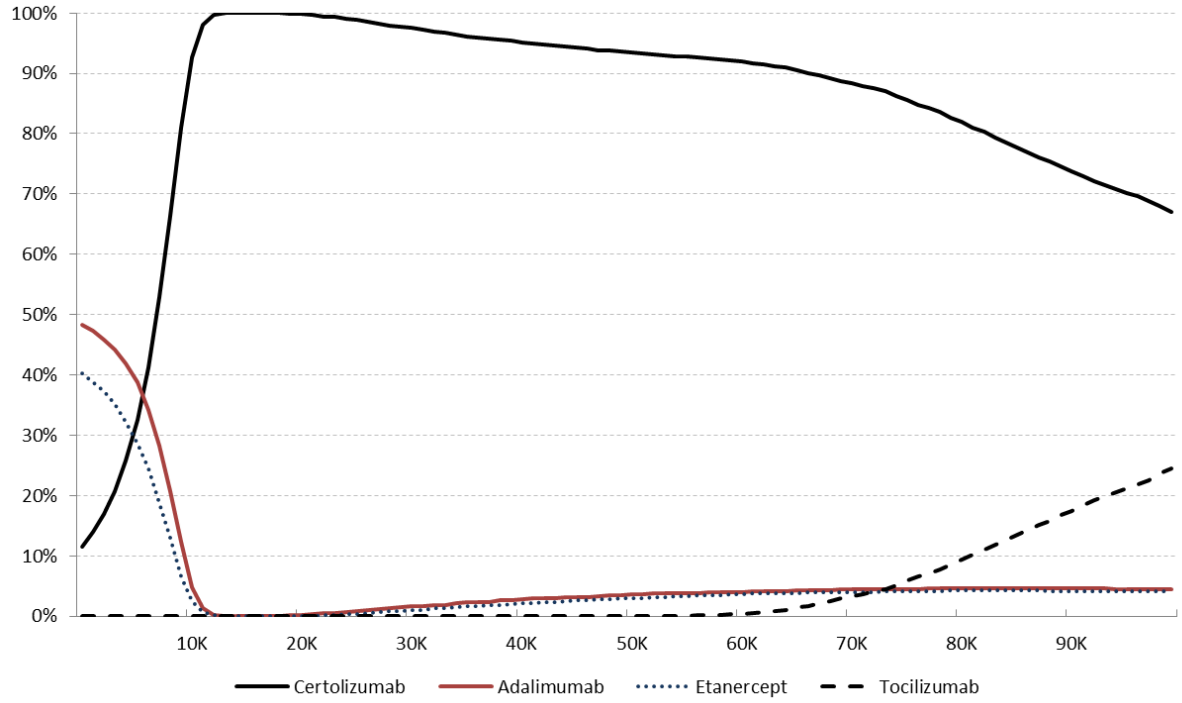


Table II: Revised probabilistic sensitivity analysis results in population C

Therapy	Mean costs	Difference in costs (CZP vs. treatment)	Mean QALYs	Difference in QALYs (CZP vs. treatment)	ICER (CZP vs. treatment)	Probability of cost effectiveness at WTP threshold of £20,000/QALY	Probability of cost effectiveness at WTP threshold of £30,000/QALY (%)
CZP	£97,254		6.162			99.78%	97.48%
ADA	£95,918	£1,336	5.902	0.259	£5,151	0.18%	1.60%
ETA	£96,270	£984	5.899	0.263	£3,746	0.04%	0.92%
TOC	£123,433	-£26,179	6.358	-0.196	£133,655	0.00%	0.00%

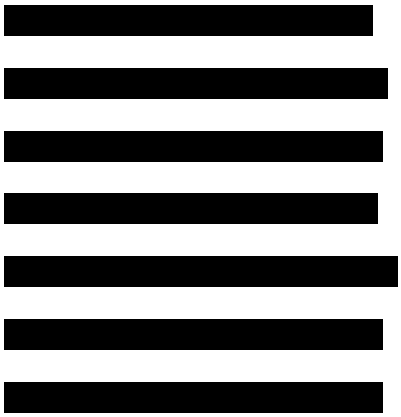
* No patient access scheme was taken into account for TOC
**Original brands only (ie Enbrel (ETA))
Results represent mean estimates per patient

Figure F: Revised Cost effectiveness acceptability curves in population C



3.3 Changes to the executable model

[redacted]



NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

Patient/carer organisation submission (STA)

Certolizumab pegol for treating rheumatoid arthritis after inadequate response to a TNF inhibitor [ID824]

Thank you for agreeing to give us your views on this treatment that is being appraised by NICE and how it could be used in the NHS. Patients, carers and patient organisations can provide a unique perspective on conditions and their treatment that is not typically available from other sources. We are interested in hearing about:

- the experience of having the condition or caring for someone with the condition
- the experience of receiving NHS care for the condition
- the experience of having specific treatments for the condition
- the outcomes of treatment that are important to patients or carers (which might differ from those measured in clinical studies, and including health-related quality of life)
- the acceptability of different treatments and how they are given
- expectations about the risks and benefits of the treatment.

To help you give your views, we have provided a questionnaire. You do not have to answer every question — the questions are there as prompts to guide you. The length of your response should not normally exceed 10 pages.

1. About you and your organisation

Your name: [redacted]

Name of your organisation: National Rheumatoid Arthritis Society

Your position in the organisation: [redacted]

Brief description of the organisation: We are the national patient organisation for and representing people with RA and JIA in the UK. We have approx 5,500 members including health professional members. We have a wide range of income streams with the majority of our funding coming from grant-giving trusts and foundations, events, legacy income and a maximum of 15% of annual income comes from projects funded by pharmaceutical industry, although to date such funding has never reached as much as 15%.

(For example: who funds the organisation? How many members does the organisation have?)

We are asking for your collective view as an organisation and will be asking patient experts for their individual input separately. If you have the condition, or care for someone with the condition, you may wish to complete a patient expert questionnaire to give your individual views as well.

Links with, or funding from the tobacco industry - please declare any direct or indirect links to, and receipt of funding from the tobacco industry: None

2. Living with the condition

What is it like to live with the condition or what do carers experience when caring for someone with the condition?

Being diagnosed with an incurable, painful disease like RA can be extremely distressing as it is life-changing and as you can be diagnosed at any age post 16, it can have a major impact on your future life plans, dreams and aspirations, although being diagnosed today has significantly better potential outcomes than when I was diagnosed 35 years ago when treatments and the way the disease was treated were quite different. RA impacts on every area of life and impacts both physical and emotional wellbeing. Health beliefs, how you come to diagnosis (how long it takes to be diagnosed), the network of support you have and how aggressive the disease is will all impact on how

you come to terms with your diagnosis and cope day to day. It can be very distressing for a partner of someone with RA to witness their loved-one in severe pain and suffering the debilitating effects of fatigue and so this disease does very much impact on the whole family. As ¾ of people are diagnosed when of working age, anxiety over job-loss due to their disease is a significant factor and whilst we are making steps towards seeing work as a health outcome, we are far from a situation where rheumatology teams pay enough attention to how worried patients may be about their job. For young people who are not yet in a permanent relationship, it can be very hard to come to terms with the fact that they have a long term condition and we know from our own research that RA can have a huge impact, making them feel less desirable, much less confident and worried that they will not find a partner. For older people diagnosed as they approach retirement for example, dreams of being able to travel and look after grand-children can suddenly seem unachievable. Diagnosed in mid-years with young children to care for can also be incredibly challenging. Imagine not being able to pick up your baby and change its nappy.

3. Current practice in treating the condition

Which treatment outcomes are important to patients or carers? (That is, what would patients or carers like treatment to achieve?) Which of these are most important? If possible, please explain why.

People simply want their life back. They want a reduction in pain, want to prevent permanent disability, want reduction in fatigue, and above all want to maintain independence and ability to work and carry out all the normal activities of daily living. Side effects of some drugs can be quite debilitating, however, by comparison to methotrexate for example, side effects from biologics are generally fewer in our experience.

What is your organisation's experience of currently available NHS care and of specific treatments for the condition? How acceptable are these treatments and which are preferred and why?

One of the key issues associated with current care is the variability of access to best, evidence-based care and access to all the relevant members of a consultant-led multi-disciplinary team. People do experience different levels of care and access to research studies for example and this can cause

confusion for patients who, particularly in the early stages of their disease, don't know what good looks like or what they should be able to ask for. This is where we come in – our goal is to be there at the start of everyone's journey and whenever they need us along the way. We try to emphasise the importance of supported self-management early on as the more you know about the disease and the more you can do to help yourself in a positive way, the better your outcomes are likely to be. Unfortunately, whilst there is a lot of rhetoric about self-management for people with LTCs, the reality of investment in it at a CCG/Trust level is few and far between. That's one of the reasons it is essential that health professionals sign-post patients to organisations who can help and support like NRAS.

4. What do patients or carers consider to be the advantages of the treatment being appraised?

Benefits of a treatment might include its effect on:

- the course and/or outcome of the condition
- physical symptoms
- pain
- level of disability
- mental health
- quality of life (such as lifestyle and work)
- other people (for example, family, friends and employers)
- ease of use (for example, tablets rather than injection)
- where the treatment has to be used (for example, at home rather than in hospital)
- any other issues not listed above

Please list the benefits that patients or carers expect to gain from using the treatment being appraised.

The key driver of RA is inflammation which can result quite quickly in bone erosion leading ultimately to joint destruction and potential disability. Many people have a positive response to a first TNF inhibitor and often for these secondary non-response patients, a second TNF option is the preferred patient/clinician choice. Cimzia at this point in the pathway therefore offers an additional therapy option which has been shown to have equal efficacy and safety profile to other TNFs available on the market and included within current NICE guidance. We still don't know who will respond to which drug

and it is clear, and I am one such patient, that you may respond to one type of Anti-TNF and not another and so there is definite need for a range of biologic therapies, even within the same class such as the Anti-TNFs, as it is impossible to predict who will respond to what with any degree of accuracy. Also as Rituximab is targeted at those with sero positive disease, those who are sero-negative (about 30%) would see an alternative TNF as preferable to moving onto RTX. When I first went onto an Anti-TNF over 15 years ago the good effects of the first Anti-TNF I was on started to wear off and stop after 3 years and so I had to move on to other biologics and am now controlled well on Humira having tried 4 Anti-TNFs as I am sero-negative and would not wish to go onto RTX. This is why we need a range of treatment options in the biologics. Biologics have enabled me to start NRAS, work 50-60 hours a week and employ 21 people. Without them I would without question be permanently in a wheelchair by now and very limited in terms of mobility and function.

Please explain any advantages that patients or carers think this treatment has over other NHS treatments in England.

Results varied from patient to patient, but in clinical trials with CIMZIA, many patients with moderate to severe rheumatoid arthritis (RA) experienced noticeable relief of their RA symptoms compared to patients taking a placebo. Noticeable RA relief: In clinical trials, CIMZIA was proven to provide noticeable RA symptom relief including improvement in pain, swelling, and tenderness for the majority of patients within 24 weeks and for some patients as fast as 1 to 2 weeks versus placebo.

Stopped further joint damage for most patients as shown by their X-rays at both 6 months and 1 year in a clinical trial versus patients in a control group.

Improvement in common everyday activities: In a clinical trial, patients on CIMZIA after 6 months reported improvement in everyday activities such as showering, getting dressed, doing chores, walking short distances, and running simple errands.

The clinical study results above were based on patients who received the recommended initial and maintenance doses of CIMZIA. These results are similar to other Anti-TNF clinical trial results but CIMZIA is slightly different to other Anti-TNFs in that it is a **PEGylated** Fab' fragment of a humanized TNF inhibitor and sometimes these small differences can mean that people will react differently to the different types of Anti-TNF drug available. Availability of Cimzia as an option in the pathway increases patient and clinical options and choice which is to be welcomed.

Rheumatologists should be able to decide on sequential prescribing. Prof Paul Emery has said "… but we have treated people who had no response after the

first anti-TNF but then went into remission with another. However, if you have a patient that has failed on two anti-TNF therapies you would be less inclined to treat with a third…."

CIMZIA delivered noticeable, proven RA relief for patients with moderate to severe RA.

If you know of any differences in opinion between patients or carers about the benefits of the treatment being appraised, please tell us about them.

Nothing to add here.

5. What do patients and/or carers consider to be the disadvantages of the treatment being appraised?

Disadvantages of a treatment might include:

- aspects of the condition that the treatment cannot help with or might make worse
- difficulties in taking or using the treatment (for example, injection rather than tablets)
- side effects (for example, type or number of problems, how often, for how long, how severe. Please describe which side effects patients might be willing to accept or tolerate and which would be difficult to accept or tolerate)
- where the treatment has to be used (for example, in hospital rather than at home)
- impact on others (for example, family, friends and employers)
- financial impact on the patient and/or their family (for example, the cost of travel to hospital or paying a carer)
- any other issues not listed above

Please list any concerns patients or carers have about current NHS treatments in England.

No disadvantages or concerns to add here. This is a sub-cut Anti-TNF which the patient can administer at home and so there are no issues around travel to hospital for treatment.

Please list any concerns patients or carers have about the treatment being appraised.

Not aware of any. Most patients want access to biologic therapies such as Cimzia whether they are eligible or not if their disease is insufficiently

controlled. Sadly we lag well behind our counterparts in Europe who are able to access biologics at an earlier stage in the pathway.

If you know of any differences in opinion between patients or carers about the disadvantages of the treatment being appraised, please tell us about them.

None

6. Patient population

Are there any groups of patients who might benefit more from the treatment than others? If so, please describe them and explain why.

We are disappointed that NICE did not uphold our joint appeal against the MTA FAD outcome and allow patients with moderate to severe disease who are failing on standard DMARD therapy and have poor prognostic markers (high CRP, erosions, ACPA +ve) but who do not have a DAS score of >5.1 to access Anti-TNF or biologic therapy but we are where we are.

Also, as mentioned earlier, people who are sero-negative may be more likely to benefit from a second TNF than going onto RTX or an alternative target biologic after failure of a first TNF thus keeping further options available for future treatment.

Are there any groups of patients who might benefit less from the treatment than others? If so, please describe them and explain why.

Not that I am aware of

7. Research evidence on patient or carer views of the treatment

Is your organisation familiar with the published research literature for the treatment?

Yes

If you answered 'no', please skip the rest of section 7 and move on to section 8.

Please comment on whether patients' experience of using the treatment as part of their routine NHS care reflects the experiences of patients in the clinical trials.

As far as we are aware

Do you think the clinical trials have captured outcomes that are important to patients? Are you aware of any limitations in how the treatment has been assessed in clinical trials?

From memory I believe that there were some interesting work-related outcomes in the original Cimzia trials. Work is an important outcome for many patients of working age.

If the treatment being appraised is already available in the NHS, are there any side effects that were not apparent in the clinical trials but have emerged during routine NHS care?

Not that I am aware of

Are you aware of any relevant research on patient or carer views of the condition or existing treatments (for example, qualitative studies, surveys and polls)?

Yes

If yes, please provide references to the relevant studies.

We have done our own research into impact on individuals and families of the condition and happy to provide access to our Family Matters report and our booklet on 'emotions, sexuality and relationships' was produced on the basis of quantitative and qualitative research undertaken. Both of these publications are downloadable from that section on our website.

8. Equality

NICE is committed to promoting equality of opportunity, eliminating unlawful discrimination and fostering good relations between people with particular protected characteristics and others. Protected characteristics are: age; being or becoming a transsexual person; being married or in a civil partnership; being pregnant or having a child; disability; race including colour, nationality, ethnic or national origin; religion, belief or lack of religion/belief; sex; sexual orientation.

Please let us know if you think that recommendations from this appraisal could have an adverse impact on any particular groups of people, such as:

- excluding from full consideration any people protected by the equality legislation who fall within the patient population for which the treatment is/will be licensed;
- having a different impact on people protected by the equality legislation than on the wider population, e.g. by making it more difficult in practice for a specific group to access the treatment;
- any adverse impact on people with a particular disability or disabilities.

Please let us know if you think that there are any potential equality issues that should be considered in this appraisal.

None I am aware of

Are there groups of patients who would have difficulties using the treatment or currently available treatments? Please tell us what evidence you think would help the Committee to identify and consider such impacts.

The only difficulty some people with longer standing disease may have is if they have significant deformity in their hands and so handling a syringe might be a challenge although the Cimzia syringe has an OXO good grips handle which does make it a little easier.

9. Other issues

Do you consider the treatment to be innovative?

Yes

If yes, please explain what makes it significantly different from other treatments for the condition.

I wouldn't describe it as significantly different as it is a TNF but it's not just another monoclonal, it is Pegylated as described above.

Are there any other issues that you would like the Appraisal Committee to consider?

10. Key messages

In no more than 5 bullet points, please summarise the key messages of your submission.

- People with seemingly similar types of disease can respond differently to the different Anti-TNFs available and we need choice and options and

adding Cimzia to the pathway post DMARD failure increases patient and clinician choice

- RA is a very nasty disease and its impact often misunderstood or minimised
- It is only tight drug control that helps to slow down or stop disease progression and sero-negative patients may benefit more from access to a second TNF than RTX.
- Uncontrolled disease can shorten lifespan
- Quality of life can be seriously impaired by RA

Appendix G - professional organisation submission template

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

Single Technology Appraisal (STA)

Certolizumab pegol for treating rheumatoid arthritis after inadequate response to a TNF inhibitor [ID824]

Thank you for agreeing to make a submission on your organisation's view of the technology and the way it should be used in the NHS.

Healthcare professionals can provide a unique perspective on the technology within the context of current clinical practice which is not typically available from the published literature.

To help you in making your submission, we have provided a template. The questions are there as prompts to guide you. It is not essential that you answer all of them.

Please do not exceed the 8-page limit.

About you

Your name: [redacted]

Name of your organisation: British Society for Rheumatology
Are you (tick all that apply):

- a specialist in the treatment of people with the condition for which NICE is considering this technology? √
- a specialist in the clinical evidence base that is to support the technology (e.g. involved in clinical trials for the technology)?
- an employee of a healthcare professional organisation that represents clinicians treating the condition for which NICE is considering the technology? If so, what is your position in the organisation where appropriate (e.g. policy officer, trustee, member etc)?
- other? (please specify)

Links with, or funding from the tobacco industry - please declare any direct or indirect links to, and receipt of funding from the tobacco industry:

none

Appendix G - professional organisation submission template

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

Single Technology Appraisal (STA)

Certolizumab pegol for treating rheumatoid arthritis after inadequate response to a TNF inhibitor [ID824]

What is the expected place of the technology in current practice?

How is the condition currently treated in the NHS? Is there significant geographical variation in current practice? Are there differences of opinion between professionals as to what current practice should be? What are the current alternatives (if any) to the technology, and what are their respective advantages and disadvantages?

Rheumatoid Arthritis (RA) is managed according to national and international guidelines (including NICE CG79[1] and European League Against Rheumatism (EULAR) guidelines[2]), aimed at controlling inflammatory disease activity as early and effectively as possible with disease-modifying drugs (DMARDs). Management of RA generally occurs within secondary care settings, which may be community-based, using an MDT approach supervised by a Consultant Rheumatologist. Patients are eligible to receive biological DMARDs (bDMARDs) if standard therapy has failed and there is an on-going high level of inflammatory disease activity (as judged by DAS28). Use of bDMARDs is according to the relevant NICE Technology Appraisal (e.g. MTA375, TA195, TA 225, TA247).

There is minimal variation in overall access to bDMARDs given the existence of NICE MTA/TAs, however there is variation in how these guidelines are interpreted and implemented. As such, several local and regional pathways have been developed between clinicians and commissioners. Currently, patients who fail to respond adequately to their initial anti-TNF therapy should next receive rituximab in combination with methotrexate; however if either are contra-indicated then a second anti-TNF can be prescribed (named as adalimumab, etanercept or infliximab). It is at this stage that inconsistency exists as certolizumab pegol is not a treatment option.

Most rheumatologists would consider the use of all available anti-TNF agents at this stage (failure of first bDMARD), and in practice all options are prescribed (including certolizumab pegol). Advantages of this technology include frequency of injection (2-weekly), monotherapy license, early response assessment (3 months vs. 6 months), and patient access scheme.

Are there any subgroups of patients with the condition who have a different prognosis from the typical patient? Are there differences in the capacity of different subgroups to benefit from or to be put at risk by the technology?

Patients with a worse prognosis tend to be identified in clinical practice by the presence of very high levels of systemic inflammation, the presence of joint damage/erosions early in the disease course (especially at presentation), the presence of rheumatoid factor and/or anti-CCP antibody, and failure of traditional therapies (especially combination therapy).

There is no currently accepted agreement that certolizumab pegol is a more or less appropriate therapeutic option in any patient group (by severity or antibody status, for example) in terms of overall efficacy or safety. However, the ability to assess

response at 3 months (rather than the normal 6 months) with certolizumab pegol may inform clinical prescribing decisions.

In what setting should/could the technology be used – for example, primary or secondary care, specialist clinics? Would there be any requirements for additional professional input (for example, community care, specialist nursing, other healthcare professionals)?

Patients receiving bDMARDs need to do so under the supervision of a Consultant Rheumatologist , who is usually based in secondary care (although may offer community-based clinical care), supported by clinical nurse specialists in rheumatology. No additional service requirements are anticipated to support this technology, were it approved.

If the technology is already available, is there variation in how it is being used in the NHS? Is it always used within its licensed indications? If not, under what circumstances does this occur?

Certolizumab pegol is already widely prescribed as an initial bDMARD within the NHS, both in combination with methotrexate and as monotherapy. It is also prescribed as a second-line anti-TNF (according to its licensed indication) in a more variable fashion, depending on local prescribing agreements.

http://gmmmg.nhs.uk/docs/guidance/GMMMG%20RA%20Pathway%2022%20april_2015.pdf

http://www.lancsmmg.nhs.uk/wp-content/uploads/sites/3/2013/04/LMMG-Recommended-Rheumatoid-Arthritis-Biologic-Pathway-September-2014.pdf

http://www.surreyandsussex.nhs.uk/wp-content/uploads/2013/04/Surrey-PCN-RA-Pathway.pdf

http://www.westhampshireccg.nhs.uk/downloads/categories/medicines/guidance/1271-nice-biologics-in-rheumatoid-arthritis-august-2015/file

Please tell us about any relevant clinical guidelines and comment on the appropriateness of the methodology used in developing the guideline and the specific evidence that underpinned the various recommendations.

RA is managed according to a variety of national and international guidelines, but the choice of biologic is most affected by NICE TAs and MTAs, and local/regional prescribing guidelines. Both NICE CG79 and EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs (2013 update)[2] utilise high-quality evidence and expert opinion to formulate treatment and prescribing advice.

[1] NICE Clinical Guideline 79. NICE. 2009.
[2] Smolen JS, Landewé R, et al. Ann Rheum Dis; doi: 10.1136/annrheumdis-2013-204573.

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

Single Technology Appraisal (STA)

Certolizumab pegol for treating rheumatoid arthritis after inadequate response to a TNF inhibitor [ID824]

The advantages and disadvantages of the technology

NICE is particularly interested in your views on how the technology, when it becomes available, will compare with current alternatives used in the UK. Will the technology be easier or more difficult to use, and are there any practical implications (for example, concomitant treatments, other additional clinical requirements, patient acceptability/ease of use or the need for additional tests) surrounding its future use?

Certolizumab pegol is already widely used on the management of RA, and the scope of this STA is to utilise this option in line with other anti-TNF agents, and within its licensed indication. As such, there are no additional practical implications or requirements to facilitate its implementation.

If appropriate, please give your view on the nature of any rules, informal or formal, for starting and stopping the use of the technology; this might include any requirements for additional testing to identify appropriate subgroups for treatment or to assess response and the potential for discontinuation.

The proposed start/stop rules will be consistent with current practice in TA195. These state that certolizumab pegol will be used in patients with an inadequate response to an initial TNF inhibitor. Unless additional evidence suggests certain subgroups will respond better, no additional testing will be required than is currently performed in routine clinical practice.

If you are familiar with the evidence base for the technology, please comment on whether the use of the technology under clinical trial conditions reflects that observed in clinical practice. Do the circumstances in which the trials were conducted reflect current UK practice, and if not, how could the results be extrapolated to a UK setting? What, in your view, are the most important outcomes, and were they measured in the trials? If surrogate measures of outcome were used, do they adequately predict long-term outcomes?

UCB refer to key trial in support of this STA, the REALISTIC study. This was published in 2012, recruiting patients from Europe and USA between 2008 and 2010, and was specifically designed to reflect patients seen in normal clinical practice. In that regard the patient population was similar to the UK, with high DAS28 at entry (mean DAS28-CRP 5.7) and longer disease duration (mean disease duration 8.6 years) than commonly seen. However, no lower limit of DAS28 was required to enter the study. Similarly, ~35% had received at least one previous TNF inhibitor.

The primary outcome of the REALISTIC study was response at 12 weeks as assessed by ACR20, which is not the usual clinical outcome used in the UK. However, DAS28 response was a pre-specified secondary outcome and patients suggested significant response rates compared to placebo. Although other outcomes were assessed at 12 weeks, such as HAQ-DI, such short term efficacy is not as

important to clinicians as longer term (>12 months) outcomes. Equally, radiographic progression data does need longer term studies to demonstrate sustained efficacy.

Overall, most clinical trial data use ACR 20 response data as a primary outcome, and the more UK-relevant DAS28 response rates are pre-specified secondary outcomes. Long term impact on HAQ and radiographic data are available for certolizumab pegol, but not universally in the patient population being assessed in this STA.

What is the relative significance of any side effects or adverse reactions? In what ways do these affect the management of the condition and the patient's quality of life? Are there any adverse effects that were not apparent in clinical trials but have come to light subsequently during routine clinical practice?

The adverse effects of the intervention in clinical trials have been consistent with those of other TNF inhibitors, and can be very significant for individual patients who suffer them. They include a higher risk of infection compared to placebo, and subsequent longer term registry studies have identified an increased risk early (<6 months) in the course of treatment. There are post-marketing and registry data to support the identification of those patients at particularly high-risk of infection, and these may be managed slightly differently (e.g RABBIT risk score). At the time of writing, no unpredicted adverse effects have come to light from observational studies or routine clinical practice, and the use of the intervention would be in line with routine clinical practice.

Any additional sources of evidence

The longer term safety and efficacy of many biologic agents, including certolizumab pegol, is being studied by various international registries including the British Society of Rheumatology Biologics Register for RA since 2010.

Implementation issues

The NHS is required by the Department of Health to provide funding and resources for medicines and treatments that have been recommended by NICE technology appraisal guidance. This provision has to be made within 3 months from the date of publication of the guidance.

There would be minimal, if any impact, on NHS resources if this STA were to be approved. This relates to the widespread use of the technology in rheumatology units for a variety of indications.

Equality
NICE is committed to promoting equality of opportunity, eliminating unlawful discrimination and fostering good relations between people with particular protected characteristics and others.

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

Single Technology Appraisal (STA)

Certolizumab pegol for treating rheumatoid arthritis after inadequate response to a TNF inhibitor [ID824]

Please tell us what evidence should be obtained to enable the Committee to identify and consider such impacts. I am not aware of any impact this appraisal would have on the populations mentioned above.

Appendix G - professional organisation submission template

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

Single Technology Appraisal (STA)

Certolizumab pegol for treating rheumatoid arthritis after inadequate response to a TNF inhibitor [ID824]

Thank you for agreeing to make a submission on your organisation's view of the technology and the way it should be used in the NHS.

Healthcare professionals can provide a unique perspective on the technology within the context of current clinical practice which is not typically available from the published literature.

To help you in making your submission, we have provided a template. The questions are there as prompts to guide you. It is not essential that you answer all of them.

Please do not exceed the 8-page limit.

About you

Your name: [redacted]

Name of your organisation: Primary Care Rheumatology Society
Are you (tick all that apply):

- a specialist in the treatment of people with the condition for which NICE is considering this technology? **YES**
- a specialist in the clinical evidence base that is to support the technology (e.g. involved in clinical trials for the technology)?
- an employee of a healthcare professional organisation that represents clinicians treating the condition for which NICE is considering the technology? If so, what is your position in the organisation where appropriate (e.g. policy officer, trustee, member etc)?
- other? (please specify)

Links with, or funding from the tobacco industry - please declare any direct or indirect links to, and receipt of funding from the tobacco industry: None

Appendix G - professional organisation submission template

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

Single Technology Appraisal (STA)

Certolizumab pegol for treating rheumatoid arthritis after inadequate response to a TNF inhibitor [ID824]

What is the expected place of the technology in current practice?

How is the condition currently treated in the NHS? Is there significant geographical variation in current practice? Are there differences of opinion between professionals as to what current practice should be? What are the current alternatives (if any) to the technology, and what are their respective advantages and disadvantages?

Are there any subgroups of patients with the condition who have a different prognosis from the typical patient? Are there differences in the capacity of different subgroups to benefit from or to be put at risk by the technology?

In what setting should/could the technology be used – for example, primary or secondary care, specialist clinics? Would there be any requirements for additional professional input (for example, community care, specialist nursing, other healthcare professionals)?

If the technology is already available, is there variation in how it is being used in the NHS? Is it always used within its licensed indications? If not, under what circumstances does this occur?

Please tell us about any relevant **clinical guidelines** and comment on the appropriateness of the methodology used in developing the guideline and the specific evidence that underpinned the various recommendations.

The body of evidence on treatment of Rheumatoid arthritis is quite conclusive on the benefits of early suppression of inflammation in patients with the disease. TNF inhibitor drugs have emerged since 1998 and provide a valuable alternative to patients who fail conventional disease modifying anti-rheumatic drugs.

Certolizumab pegol is a fragment of a humanized monoclonal antibody that is unique from other TNF inhibitors in that it lacks the constant fragment of Immunoglobulin and therefore is less likely to cause antibody dependent or complement dependent cytotoxicity. It may also convey less potential for Immunogenicity. It is currently used in severe DMARD resistant rheumatoid arthritis and can be used in combination with Methotrexate or as a monotherapy.

It is currently in use in the NHS but its availability and use is varied. One local Midlands protocol suggests that Certolizumab should only be used after failed DMARD treatments and after a trial with Rituximab. This appears to be a local policy rather than standard practice. There is variation across the country about which TNF inhibitors should be used first line as several alternatives exist. There should be greater clarity in this area.

We feel the use of Certolizumab pegol should be restricted to secondary care rheumatology clinics due to the cost and the skills required in its administration and the necessary screening for contraindications that needs to take place prior to administering the drug.

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

Single Technology Appraisal (STA)

Certolizumab pegol for treating rheumatoid arthritis after inadequate response to a TNF inhibitor [ID824]

The advantages and disadvantages of the technology

NICE is particularly interested in your views on how the technology, when it becomes available, will compare with current alternatives used in the UK. Will the technology be easier or more difficult to use, and are there any practical implications (for example, concomitant treatments, other additional clinical requirements, patient acceptability/ease of use or the need for additional tests) surrounding its future use?

If appropriate, please give your view on the nature of any rules, informal or formal, for starting and stopping the use of the technology; this might include any requirements for additional testing to identify appropriate subgroups for treatment or to assess response and the potential for discontinuation.

If you are familiar with the evidence base for the technology, please comment on whether the use of the technology under clinical trial conditions reflects that observed in clinical practice. Do the circumstances in which the trials were conducted reflect current UK practice, and if not, how could the results be extrapolated to a UK setting? What, in your view, are the most important outcomes, and were they measured in the trials? If surrogate measures of outcome were used, do they adequately predict long-term outcomes?

What is the relative significance of any side effects or adverse reactions? In what ways do these affect the management of the condition and the patient's quality of life? Are there any adverse effects that were not apparent in clinical trials but have come to light subsequently during routine clinical practice?

This technology is used in similar ways to other TNF inhibitor agents and is given either subcutaneously or intravenously. The safety profile appears similar to other TNF inhibitor drugs. Adverse events from Certolizumab alone do not appear to differ from a combination of Certolizumab and Methotrexate combination. It appears to have lower incidence of injection site reactions and injection site pain, thereby could be more acceptable to patients.

There is already experience of using Certolizumab in practice with effectiveness that is comparative to other TNF inhibitors in practice. It can be used in combination with Methotrexate or as a monotherapy if methotrexate is not tolerated.

There is an increased risk of infection with all TNF inhibitor drugs and Certolizumab appears comparable in terms of adverse events with other TNF inhibitors. The most common infections tend to be respiratory tract infections and also Tuberculosis, but this is the case for other TNF inhibitors as well.

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

Single Technology Appraisal (STA)

Certolizumab pegol for treating rheumatoid arthritis after inadequate response to a TNF inhibitor [ID824]

Any additional sources of evidence

Can you provide information about any relevant evidence that might not be found by a technology-focused systematic review of the available trial evidence? This could be information on recent and informal unpublished evidence, or information from registries and other nationally coordinated clinical audits. Any such information must include sufficient detail to allow a judgement to be made as to the quality of the evidence and to allow potential sources of bias to be determined.

None

Implementation issues

The NHS is required by the Department of Health to provide funding and resources for medicines and treatments that have been recommended by NICE technology appraisal guidance. This provision has to be made within 3 months from the date of publication of the guidance.

If the technology is unlikely to be available in sufficient quantity, or the staff and facilities to fulfil the general nature of the guidance cannot be put in place within 3 months, NICE may advise the Department of Health to vary this direction.

Please note that NICE cannot suggest such a variation on the basis of budgetary constraints alone.

How would possible NICE guidance on this technology affect the delivery of care for patients with this condition? Would NHS staff need extra education and training? Would any additional resources be required (for example, facilities or equipment)?

The main limitation of Certolizumab is the cost of the drug. This is the same with all other TNF inhibitor drugs. Its cost effectiveness seems to be accepted in patients who have failed initial DMARD therapy but there is no evidence to suggest its effectiveness in early disease before the failure of conventional DMARD treatments.

The development of biomarkers that could suggest individual patients' response to therapy would be invaluable in helping clinicians with their choice of TNF inhibitor treatements.

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

Single Technology Appraisal (STA)

Certolizumab pegol for treating rheumatoid arthritis after inadequate response to a TNF inhibitor [ID824]

Equality
NICE is committed to promoting equality of opportunity, eliminating unlawful discrimination and fostering good relations between people with particular protected characteristics and others. Please let us know if you think that this appraisal:

- could exclude from full consideration any people protected by the equality legislation who fall within the patient population for which [the treatment(s)] is/are/will be licensed;
- could lead to recommendations that have a different impact on people protected by the equality legislation than on the wider population, e.g. by making it more difficult in practice for a specific group to access the technology;
- could lead to recommendations that have any adverse impact on people with a particular disability or disabilities.

Please tell us what evidence should be obtained to enable the Committee to identify and consider such impacts.

None known

Certolizumab pegol for treating rheumatoid arthritis after inadequate response to a TNF inhibitor: A Single Technology Appraisal

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

Authors Matt Stevenson, Professor of Health Technology Assessment, ScHARR, University of Sheffield, Sheffield, UK

Inigo Bermejo, Research Associate, ScHARR, University of Sheffield, Sheffield, UK

Rachel Archer, Research Fellow, ScHARR, University of Sheffield, Sheffield, UK

John Stevens, Reader in Decision Science, ScHARR, University of Sheffield, Sheffield, UK

Edward Goka, Research Associate, ScHARR, University of Sheffield, Sheffield, UK

Mark Clowes, Information Specialist, ScHARR, University of Sheffield, Sheffield, UK

David Scott, Department of Rheumatology, King's College Hospital NHS Foundation Trust, London, UK

Adam Young, Department of Rheumatology, West Hertfordshire Hospital NHS Trust, Hertfordshire, UK

Correspondence Author Matt Stevenson, Professor of Health Technology Assessment, ScHARR, University of Sheffield, Sheffield, UK

Date completed Date completed (17/05/2016)

Source of funding: This report was commissioned by the NIHR HTA Programme as project number 15/06/07.

Declared competing interests of the authors

John Stevens undertook methodological work for BMS, the manufacturer of abatacept, for ScHARR, although this work was not related to rheumatoid arthritis. The remaining authors have no conflicts of interest to declare.

Acknowledgements

We would like to thank Paul Tappenden, ScHARR, for providing comments on the draft report and Andrea Shippam, Programme Manager, ScHARR, for providing administrative support and in preparing and formatting the report.

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:

Stevenson M, Bermejo I, Archer R, Stevens J, Goka E, Clowes M, Scott D, Young A. Certolizumab pegol for treating rheumatoid arthritis after inadequate response to a TNF inhibitor: A Single Technology Appraisal. School of Health and Related Research (ScHARR), 2016.

Contributions of authors

Rachel Archer and Edward Goka summarised and critiqued the clinical effectiveness data reported within the company's submission. Inigo Bermejo and Matt Stevenson critiqued the health economic analysis submitted by the company. John Stevens critiqued the network meta-analysis and provided other statistical support. Mark Clowes critiqued the company's search strategy. David Scott and Adam Young provided clinical advice to the team. All authors were involved in drafting and commenting on the final report.

CONTENTS

List of tables

List of figures

Abbreviations

ABA	abatacept
AC	Appraisal Committee
ACR	American College of Rheumatology
ACR20	20% improvement in the ACR score
ACR50	50% improvement in the ACR score
ACR70	70% improvement in the ACR score
ADA	adalimumab
AE	adverse event
AIC	Akaike Information Criterion
AZA	azathioprine
bDMARD	biologic disease-modifying antirheumatic drug
BNF	British National Formulary
BSRBR	British Society for Rheumatology Biologics Register
cDMARD	conventional disease-modifying antirheumatic drug
CDAI	Clinical Disease Activity Index
CI	confidence interval
CRD	Centre for Reviews and Dissemination
CrI	credible interval
CS	company's submission
DAS28	Disease Activity Score 28
DMARD	disease-modifying antirheumatic drug
EQ-5D	EuroQol 5 Dimensions
ERAS	Early RA Study
ERG	Evidence Review Group
ESR	Erythrocyte Sedimentation Rate
ETA	etanercept
EULAR	European League Against Rheumatism
FAD	Final Appraisal Determination
GLD	gold injections
GOL	golimumab
HAQ	Health Assessment Questionnaire
HAQ-DI	Health Assessment Questionnaire disability index
HCQ	hydroxychloroquine
HR	hazard ratio
HRQoL	health-related quality of life

ICER	incremental cost effectiveness ratio
IFX	infliximab
IPS	individual patient simulation
ITT	intention to treat
IV	intravenous
LOCF	last observation carried forward
MD-HAQ	Multidimensional HAQ
MTA	Multiple Technology Appraisal
mTSS	modified Total Sharp Score
MTX	methotrexate
NBT	non-biologic treatment
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
NMA	network meta-analysis
NOAR	Norfolk Arthritis Register
NR	not reported
NRI	non-responder imputation
OLE	open-label extension
ONS	Office for National Statistics
PAS	Patient Access Scheme
PBO	placebo
PSS	Personal Social Services
PSSRU	Personal Social Services Research Unit
QALY	quality-adjusted life year
Q2W	Every two weeks
Q4W	Every four weeks
RA	rheumatoid arthritis
RCT	randomised controlled trial
RF	rheumatoid factor
RR	rate ratio
RTX	rituximab
SAE	serious adverse event
SC	subcutaneous
SD	standard deviation
SE	standard error
SF-36	Short Form (36) Health Survey

SmPC	summary of product characteristics
STA	Single Technology Appraisal
SSZ	sulfasalazine
SW28	Swelling 28 joints
TEN28	Tenderness 28 joints
TNF	tumour necrosis factors
TNFi	tumour necrosis factors inhibitor
TNFi-IR	TNFi inadequate response
TOC	tocilizumab

1 SUMMARY

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

The Evidence Review Group (ERG) considers the company's description of the underlying health problem in the company's submission (CS) to be appropriate, mostly up to date and relevant to the decision problem in the final National Institute for Health and Care Excellence (NICE) scope. However, a description of rheumatoid arthritis (RA) response criteria was lacking.

The CS generally adhered to the NICE scope. One exception was the exclusion of best supportive care as a comparator, which the company omitted primarily as this 'does not reflect the current NICE recommendations for tumour necrosis factors inhibitors (TNFis)'. The clinical advisors to the ERG agreed that this was the case. A further possible deviation was in expanding the moderate to severe definition to include any patient with a Disease Activity Score 28 (DAS28) score in excess of 3.2. This contrasts with the definition in the recent multiple technology appraisal for biologic disease-modifying antirheumatic drugs (bDMARDs). However, the ERG believes that the approach taken within the CS is logical, but comments that results divided into those with a DAS28 score of >3.2 and ≤5.1. and those with a DAS28 score of >5.1 would have been beneficial as bDMARDs are only recommended by NICE in the latter group. The majority of outcomes reported in the final NICE scope have been included in the CS. Three defined outcome measures have, however, been excluded. These are: joint damage; radiological progression; and extra-articular manifestations of the disease.

1.2 Summary of clinical effectiveness evidence submitted by the company

The clinical effectiveness evidence for certolizumab pegol (CZP) in the treatment of moderate to severe RA in patients with a previously inadequate response or intolerance to TNFi therapy was based on six key randomised controlled trials (RCTs) (REALISTIC, DOSEFLEX, PREDICT, SWITCH, J-RAPID and HIKARI). All of these trials recruited both TNFi-naïve and TNFi-experienced patients, with the exception of the SWITCH study, which was performed in a TNFi-experienced population. Evidence was presented in the CS for the efficacy of CZP in combination with methotrexate (MTX) / conventional disease-modifying antirheumatic drugs or as monotherapy. Five RCTs were placebo (PBO)-controlled (PREDICT did not have a non-CZP comparator arm). The durations of the randomised controlled phases in the RCTs were 12 weeks (REALISTIC and SWITCH), 16 weeks (DOSEFLEX), 24 weeks (J-RAPID and HIKARI) and PREDICT (52 weeks). The DOSEFLEX trial had an open-label run-in phase during which all participants received CZP before responding patients were randomised into the PBO-controlled period of the study. The primary outcome in four of the included RCTs (REALISTIC, SWITCH, J-RAPID and HIKARI) was ACR20 response at week 12. The primary endpoint of DOSEFLEX was ACR20 response at 34 weeks in patients randomised at week 18. The PREDICT study was designed to compare the use of two assessment tools (clinical

disease activity scale (CDAI) and RAPID-3) and so the primary endpoint was scores on these two measures at 12 and 52 weeks. J-RAPID and HIKARI were performed exclusively in Japan. All six RCTs were judged by the company to be of good quality. The company also included supplementary observational evidence from the Swedish registry-based study ARTIS. Safety evidence was summarised from the included CZP RCTs, a pooled analysis of CZP safety and three additional safety studies (Yun *et al.* 2016; Simard *et al.* 2011, Curtis *et al.* 2015). No head to head evidence evaluating CZP against comparator bDMARDs was identified and therefore network meta-analyses were presented in the CS.

Disease activity was reported in the CS as ACR and EULAR responses, DAS28 and CDAI. The summary of ACR and EULAR response data (as key outcomes) was prioritised in the ERG report.

ACR response data were available from all included CZP RCTs, with PREDICT reporting modified ACR (mACR). By the end of the PBO-controlled phase of REALISTIC (week 12), TNFi-experienced patients who received CZP + MTX were more to be ACR20 ACR50 and ACR70 responders compared with those on PBO ACR responses were maintained to week 28 of the CZP OLE. TNFi-experienced patients treated with CZP in DOSEFLEX, J-RAPID, HIKARI and SWITCH achieved more favourable ACR responses at the end of their controlled phases compared with PBO groups. Participants in PREDICT who received CZP experienced benefits in terms of mACR response between weeks 2 and 52. Findings from the classical meta-analysis Data for ACR response in TNFi-experienced patients receiving CZP as monotherapy were available for two PBO-controlled CZP RCTs (REALISTIC and HIKARI). proportions of TNFi-experienced subjects in the REALISTIC receiving CZP monotherapy reached ACR20 and ACR70 responses by week 12 versus PBO Similar proportions of CZP and PBO participants reached an ACR50 response. Patients in HIKARI who were TNFi-experienced and treated with CZP monotherapy were more be ACR20, ACR50 and ACR70 responders at weeks 12 and 24 compared with PBO. mACR responses among TNF-experienced patients receiving CZP monotherapy and MTX combination treatment in PREDICT and appeared to be between the groups. Results from the classical meta-analysis indicated a more favourable ACR response for CZP monotherapy patients compared with PBO, although the results were inconclusive.

EULAR response data were available for all included CZP RCTs. TNFi-experienced patients receiving CZP + MTX in REALISTIC were more to achieve good or moderate EULAR

responses [redacted] at week 12 than those on PBO [redacted] The proportion of patients who achieved a good EULAR response [redacted] for patients receiving CZP + MTX than CZP monotherapy [redacted] EULAR responses were also observed for CZP + MTX-treated patients in DOSEFLEX and J-RAPID versus PBO. Whilst CZP-treated patients in the SWITCH trial had a more favourable EULAR response at week 12, results at week 24 were broadly comparable between CZP and PBO treatment groups. TNFi-experienced subjects in REALISTIC who received CZP as monotherapy experienced [redacted] of EULAR good response compared with PBO at 12 weeks [redacted] proportion of those treated with CZP monotherapy had a EULAR moderate response [redacted] %). In HIKARI, more CZP monotherapy-treated TNFi-experienced patients [redacted] or moderate [redacted] CZP benefits [redacted] at 24 weeks for good EULAR response [redacted] similar between groups for moderate EULAR response [redacted] EULAR data at 3 months were meta-analysed to compare the effects of CZP (both in combination with MTX and as monotherapy) with PBO. As previously for ACR response, the forest plots [redacted] for the intervention groups in terms of EULAR response, [redacted]

Four RCTs (REALISTIC, DOSEFLEX, PREDICT and SWITCH) presented Health Assessment Questionnaire Disability Index (HAQ-DI) scores (or MD-HAQ/M-HAQ in the case of PREDICT). CZP-treated patients experienced [redacted] in HAQ-DI relative to PBO [redacted] No data for the outcomes of joint damage or radiological progression were included in the CS. Patients receiving CZP demonstrated [redacted] fatigue in the REALISTIC and DOSEFLEX trials. The CZP group in DOSEFLEX experienced [redacted] to health-related quality of life, with [redacted] Short Form (36) Health Survey scores

Data from the ARTIS registry study were included in the CS and showed significant benefits of CZP treatment in TNFi-experienced patients in DAS28 ($p<0.0001$) and HAQ ($p<0.0001$) at 3 and 6 months following initiation of CZP treatment.

More TNFi-experienced patients in REALISTIC receiving CZP reported an AE (68.1%) than PBO (50.0%). A greater proportion of CZP-treated patients (59.3%) in SWITCH had an AE versus 40.0% on PBO. For TNFi-experienced REALISTIC patients, slightly more CZP-treated patients (7.9%) reported SAEs than those on PBO (5.0%). In DOSEFLEX, CZP treatment groups were more likely to have SAEs (CZP 200 mg every two weeks 7.1%, CZP 400 mg every 4 weeks 2.9%) than on PBO

(0.0%). A greater proportion of TNFi-experienced patients in REALISTIC who received CZP 200 mg reported infections/infestations (29.3%) compared with the PBO group (23.8%).

The results from the Bayesian network meta-analysis (NMA) within the CS indicated that: for population A [redacted]; for population B [redacted]; and for population C [redacted].

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

The ERG was satisfied that the searches for clinical effectiveness evidence reported in the CS were likely to have identified all relevant published RCT evidence. The eligibility criteria applied in the selection of evidence for the clinical effectiveness review were considered by the ERG to be reasonable and generally consistent with the decision problem as outlined in the final NICE scope. The clinical advisors to the ERG did not highlight any additional relevant RCTs that should have been included in the CS. A CZP RCT by Kang *et al.* (2012) was identified by the ERG and clarification sought from the company as to why it was not included in the CS. The company responded that the Kang trial was not included in the CS because only low numbers of patients in the trial were TNFi-experienced. However, the ERG noted that two CZP RCTs were included in the CS that had low numbers of TNFi-experienced patients (J-RAPID and HIKARI) and therefore the ERG considered that additional justification should have been provided by the company to support their decision to exclude the Kang trial.

The quality of the included CZP RCTs and ARTIS non-randomised study were assessed using well established and recognised criteria. Data for radiological progression and joint damage were not presented in the CS, however, data on inhibition of joint structural damage were available in the published articles for both J-RAPID and HIKARI. Extra-articular manifestations of disease were not included in the CS. Study and patient characteristics for included CZP trials were clearly described in a narrative summary alongside clinical and safety data. However, *p*-values were frequently unreported in the CS and therefore the ERG requested that these be provided by the company where available. Classical meta-analyses were performed for CZP used in combination with MTX and for CZP as monotherapy. Classical meta-analyses were performed separately for the outcomes of ACR20/50/70; EULAR response; and DAS28(Erythrocyte Sedimentation Rate (ESR)) remission at 3 months. No meta-analysis was performed for outcomes at 6 months due to data unavailability. Both fixed effects (Mantel-Haenszel) and random effects (DerSimonian and Laird) models were used. Heterogeneity between trials was investigated using I^2 values. The ERG noted that it is generally recommended that at least five studies should be available for a classical meta-analysis, whereas the analyses in the CS

included, at most, only three studies. A Bayesian NMA was performed to assess CZP against comparator interventions.

The ERG believes that there were several limitations with the NMA as presented within the CS. Several changes are required to the analyses conducted and to the reporting of the results in order for them to represent genuine uncertainty and be useful for decision-making purposes, including: incorporating weakly informative prior information for the between-study standard deviation; generating predictive distributions of the effects of treatments in a new study; using the evidence from the REALISTIC study to generate the probabilities of being in each ACR and EULAR category for the reference treatment; and taking draws from the joint posterior distribution of treatment effects rather than assuming univariate normal distributions for them. It was not possible for the ERG in the time available to make the required changes to produce robust results and the ERG has not amended the NMA presented in the CS.

1.4 Summary of cost effectiveness submitted evidence by the company

The manufacturer supplied a *de novo* cohort Markov model constructed in Microsoft Excel©. The perspective used was that of the NHS in England. The cycle length was set to six months and a lifetime time horizon (45 years) was used. A discount rate of 3.5% per annum was used both for costs and utilities. Patients enter the model through the initial state after inadequate response to a TNFi and transition to one of three states depending on their EULAR response: none, moderate or good. Non-responders discontinue treatment after a cycle and transition to the state representing the first six months of the first follow-up treatment. Good and moderate EULAR responders remain in their states until treatment discontinuation, when they transition to the state representing the first six months of the first follow-up treatment. Patients achieving good or moderate EULAR response in follow-up treatments transition to a state representing the rest of the duration of the treatment; non-responders transition to the state representing the first six months of the next follow-up treatment in the sequence. During any cycle, patients can transition from any of the alive states to death. Utilities depend on the EULAR response and the HAQ score at each cycle. HAQ score depends on both EULAR response and the type of treatment (bDMARD or cDMARD). Mortality is also assumed to be affected by the HAQ score.

The company considered three different populations: Population A, formed by patients eligible for rituximab in combination with methotrexate (RTX + MTX); Population B, formed by patients for whom RTX is contraindicated or withdrawn due to an adverse event; and Population C, formed by patients for whom MTX is contraindicated or withdrawn due to an adverse event. For Population A, the company compared a sequence that it believed to reflect current recommendations and clinical practice with a sequence consisting of a treatment of CZP in combination with MTX (CZP + MTX)

followed by the comparator sequence. For Population B, the company compared a sequence starting with a treatment of CZP + MTX with the sequences starting with treatments of adalimumab (ADA), golimumab (GOL), etanercept (ETA), infliximab (IFX), abatacept (ABA), and tocilizumab (TOC) each in combination with MTX. For Population C, the company compared a sequence starting with a treatment of CZP mono therapy with sequences starting with treatments of ADA, ETA and TOC monotherapies. Effectiveness data for CZP were derived from the TNFi experienced subgroup of the REALISTIC trial. The company conducted a network meta-analysis (NMA) to estimate the effectiveness of CZP and its comparators in combination with MTX.

Unit costs were taken from the Personal Social Services Research Unit, British National Formulary (BNF), and NHS Reference Costs. The cost of CZP and GOL used in the model included the public Patient Access Scheme (PAS) in place. The list prices reported in the BNF were used for the rest of the drugs, as directed by NICE, although a commercial-in-confidence PAS is in place for both ABA and TOC.

In their base case analysis, the company estimates that for Population A, the probabilistic incremental cost-effectiveness ratio (ICER) of prepending a treatment CZP + MTX to the currently recommended treatment sequence is £33,222 per quality-adjusted life year (QALY) gained (0.290 QALYs gained at a cost of £9,842). For Population B, the estimated probabilistic ICER of CZP + MTX versus GOL + MTX is £3,461 (0.256 QALYs gained at a cost of £884) whilst the estimated probabilistic ICER of TOC (intravenous (IV)) + MTX versus CZP + MTX is £132,783 (0.201 QALYs gained at a cost of £26,659). For Population C, the estimated probabilistic ICER of CZP monotherapy versus ADA is £3,461 (0.260 QALYs gained at a cost of £1,336) whilst the estimated probabilistic ICER of TOC (IV) monotherapy versus CZP monotherapy is £133,655 (0.196 QALYs gained at a cost of £26,179). One-way sensitivity analyses undertaken by the company, where the mean values were replaced with values from the relevant 95% confidence intervals, show that the net monetary benefit of CZP, assuming a threshold of £30,000 per QALY gained is most sensitive to the efficacies of rituximab (RTX) + MTX, CZP (as monotherapy or in combination with MTX), and TOC (as monotherapy or in combination with MTX). Scenario analyses undertaken by the company show that assuming the efficacy of CZP is equal to the other TNFis has the biggest impact on the ICER, followed by the treatment duration of RTX + MTX and assuming a flat HAQ score progression for cDMARDs and palliative care.

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

The ERG has concerns regarding the NMAs used to estimate the efficacy of CZP and its comparators used to characterise uncertainty in the economic model. Heterogeneity was expected but the company assumed that the treatment effects were constant across studies. The evidence for the reference

treatment from the REALISTIC study was assumed by the company to represent the evidence for the target population; however, the company only used the "no response" rates from the REALISTIC study and used evidence from all other studies to estimate the response rates for other ACR and EULAR response rates. The company generated approximate estimates of absolute probabilities being in each ACR and EULAR category by not using the results of the NMA appropriately to estimate them.

The ERG believes that the treatment sequences compared for Population A are inappropriate because they include TOC + MTX followed by ABA + MTX after RTX + MTX. Clinical experts consulted by the ERG claimed that usually TOC + MTX or ABA + MTX were provided, but not both.

The ERG notes that the company did not identify evidence on the efficacy of IFX, ADA and ETA in combination with MTX in patients with inadequate response to a TNFi. Similarly, the ERG notes that the company did not identify evidence on the efficacy of TOC, ADA and ETA monotherapies in patients with inadequate response to a TNFi. The ERG believes that the assumptions made by the company to overcome the lack of evidence did not evaluate fully the uncertainty and that therefore the outcomes of the model should be interpreted with caution.

The ERG believes that the methodology for modelling first and subsequent treatments is limited and can result in implausible sequences when comparing elongated sequences, such as the intervention sequence in Population A with shorter sequences (the comparator in Population A).

The ERG has concerns regarding the modelling of the efficacy of subsequent treatments due to the lack of evidence on treatment efficacy in patients with an inadequate response to a previous TNFi.

The ERG notes that the company assumed the same treatment duration for all bDMARDs, despite evidence suggesting different treatment durations for different bDMARDs. The ERG notes that the company identified treatment duration as a parameter with a large impact on the ICER (especially in Population A) in one of their scenario analyses.

The ERG notes that the company used a rather simple approach to map changes in HAQ score to changes in EQ-5D utility and better approaches exist to capture the non-linearity of the relationship between HAQ score and EQ-5D.

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

1.6.1 Strengths

The ERG were satisfied that the key RCTs for the efficacy of CZP in patients with an inadequate response or intolerance to TNFis were included in the CS. The included CZP RCTs were considered by both the company and the ERG to be of good quality.

The model used appears conceptually appropriate with only a few minor implementation errors, some of which were fixed during the clarification process. It contained the functionality to assess the impact of changing parameters and relevant structural uncertainties on the ICER. A number of built-in alternative scenarios were included.

1.6.2 Weaknesses and areas of uncertainty

The ERG considers that additional justification for the omission of the Kang trial from the CS should have been provided by the company. Whilst the REALISTIC trial contained the largest TNFi-experienced population, some of the findings for individual trials in the clinical effectiveness review were based on small sample sizes.

The company performed two sets of NMAs: 1) analyses of the individual categories of ACR and EULAR, 2) separate analyses of the ACR and EULAR data by assuming that the ACR and EULAR categories formed single sets of ordinal data, respectively. In both cases, the company's preferred results assumed that the treatment effects were constant across studies ignoring potential heterogeneity. Absolute probabilities of being in each ACR and EULAR category ignored the underlying joint distribution between parameters generated by the second set of NMAs, underestimated uncertainty and used evidence from all studies to partition responders between different categories rather than using evidence only from the REALISTIC study.

A further area of uncertainty is the efficacy of ADA, IFX, ETA in combination with MTX and of TOC, ADA and ETA monotherapies in patients who have had an inadequate response to a TNFi.

The company did not identify relevant evidence and assumed that the efficacy of ADA, IFX and ETA in combination with MTX was equal to that of GOL + MTX and that the relative efficacy of TOC, ADA and ETA monotherapies compared with CZP monotherapy would be the same as that when these interventions were used in combination with MTX.

The absence of evidence on the efficacy of bDMARDs in patients who have had an inadequate response to two or more bDMARDs introduced considerable uncertainty in the model. The company adopted simplifying assumptions in order to model the efficacy of bDMARDs in subsequent

treatments, the impact of which is unknown, although results comparing elongated and standard sequences appeared implausible. As such, the ERG believes that the credibility of comparisons of sequences of different lengths within the model is limited.

The company acknowledged limitations inherent to the modelling approach used when trying to model features of the disease. In particular, the company claimed that: a non-linear mapping from HAQ score to EQ-5D could not be properly applied within a cohort Markov model; and, that time-dependent transitions (for example those from a Weibull distribution) could not be implemented in states other than the initial state.

1.7 Summary of exploratory and sensitivity analyses undertaken by the ERG

The ERG applied a series of modifications to the company's base case analysis. The most relevant were: (i) adding biosimilars of IFX and ETA and subcutaneous (SC) formulations of TOC and ABA as comparators; (ii) adding two sequences to be evaluated in Population A (iii) removal of ABA + MTX treatment after TOC + MTX from the compared sequences in Population A; (iv) using different durations for different treatments based on the data provided in TA195; (v) setting the RTX retreatment interval to 7.35 months; (vi) using the results of the NMA including J-RAPID; (vi) considering all doses of TOC and setting the 800mg limit per administration recommended in TOC's summary of product characteristics; and (vii) adjusting the mean HAQ improvements reported in RADIATE to be more appropriate for responders.

These modifications resulted in the sequence including CZP + MTX being dominated in population A in the ERG's base case analysis. For population B, the estimated probabilistic ICER of CZP + MTX versus GOL + MTX is £13,155 (0.287 QALYs gained at a cost of £3,774) whilst the estimated probabilistic ICER of TOC (SC) + MTX versus CZP + MTX is £43,994 (0.544 QALYs gained at a cost of £23,954). For population C, the estimated probabilistic ICER of CZP monotherapy versus ADA is £14,437 (0.291 QALYs gained at a cost of £4,206) whilst the estimated probabilistic ICER of TOC (SC) monotherapy versus CZP monotherapy is £45,090 (0.525 QALYs gained at a cost of £23,690).

The ERG also undertook two scenario analyses: using the results from the NMA excluding J-RAPID, as the company did for its base case; and assuming ADA, ETA and IFX had the same efficacy as CZP (instead of assuming their efficacy is equal to that of GOL). The first scenario analysis has little impact on the results; contrastingly, the second scenario analysis shows very different results in which the ETA biosimilar dominates CZP in Populations B and C (Population A is unaffected). However, there remain treatments currently recommended by NICE that CZP is estimated to be more cost-effective than.

Estimates of the cost-effectiveness of CZP when the ABA and TOC PASs are taken into consideration are provided in a confidential appendix.

The conclusions of the ERG should tally with the expectations before constructing a mathematical model. The CS suggests that CZP + MTX and RTX + MTX [redacted] RTX is noticeably less expensive than CZP [redacted] CZP + MTX is estimated to b [redacted] when compared with those treatments recommended for Population B [redacted] would be logical that CZP + MTX [redacted]

2 BACKGROUND

2.1 Critique of company's description of underlying health problem

The Evidence Review Group (ERG) considers the company's description of the underlying health problem in the company's submission (CS) to be appropriate, mostly up to date and relevant to the decision problem in the final NICE scope. However, a description of rheumatoid arthritis (RA) response criteria was lacking. The ERG provides it along with a summary of the underlying health problem below.

Clinical features of rheumatoid arthritis

RA is a chronic inflammatory disease characterised by progressive, irreversible, joint damage, impaired joint function, pain and tenderness caused by swelling of the synovial lining of joints and is manifested with increasing disability and reduced quality of life.[1] The primary symptoms are pain, morning stiffness, swelling, tenderness, loss of movement, fatigue, and redness of the peripheral joints.[2, 3] RA is associated with substantial costs both directly (associated with drug acquisition and hospitalisation) and indirectly due to reduced productivity.[4] RA has long been reported as being associated with increased mortality,[5, 6] particularly due to cardiovascular events.[7]

Epidemiology

There are an estimated 580,000 people in England and Wales with RA, with approximately 26,000 incident cases per year.[8] RA is more prevalent in females (1.16%) than in males (0.44%)[9] with the majority of cases being diagnosed between the ages of 40 and 80.[10]

Aetiology

There is no identified specific cause for RA, but there seems to be a variety of contributing factors such as genetic and environmental influences. Genetic factors have a substantial contribution to RA: the heritability of RA is estimated to be between 53 and 65%[11] and a family history of RA is related with a risk ratio of 1.6 compared with the general population.[12] Many genes associated with susceptibility to RA are concerned with immune regulation. Infectious agents have been suspected but no consistent relationship with an infective agent has been proven. Similarly, sex hormones have been suspected due to the higher prevalence of RA in women and a tendency for the disease to improve during pregnancy. However, a precise relationship has not been identified. There is no proof of any causal link with lifestyle factors such as diet, smoking, or occupation.

Management of rheumatoid arthritis

Traditionally, patients have been treated with conventional disease-modifying anti-rheumatic drugs (cDMARDs) which include methotrexate (MTX), sulfasalazine (SSZ), hydroxychloroquine (HCQ),

leflunomide (LEF), and gold injections (GLD) as well as corticosteroids, analgesics and non-steroidal anti-inflammatory drugs (NSAIDs). However, more recently, a group of biologic immunosuppressant drugs have been developed that specifically modify the disease process by blocking key protein messenger molecules (such as cytokines) or cells (such as B-lymphocytes).[13] Such drugs have been labelled as biologic disease-modifying anti-rheumatic drugs (bDMARDs): certolizumab pegol (CZP), adalimumab (ADA), etanercept (ETA), golimumab (GOL), and infliximab (IFX) are tumour necrosis factor (TNF) inhibitors (or antagonists) (TNFi). Of the remaining bDMARDs, tocilizumab (TOC) is a cytokine interleukin-6 inhibitor, abatacept (ABA) is a selective modulator of the T lymphocyte activation pathway, and rituximab (RTX) is a monoclonal antibody against the CD20 protein.

Assessment of response to therapy

The initial response criteria for RA were produced in 1987 by the American College of Rheumatology[14] (ACR). NICE Clinical Guideline (CG) 79 provides a summary of the ACR criteria, namely that patients must have at least four of seven criteria: morning stiffness lasting at least 1 hour; swelling in three or more joints; swelling in hand joints; symmetric joint swelling; erosions or decalcification on x-ray of hand; rheumatoid nodules; and abnormal serum rheumatoid factor. For the first four criteria, these must have been present for a period of at least six weeks. However, in NICE Clinical Guideline 79 the guideline development group preferred a clinical diagnosis of RA rather than the ACR criteria because 'an early persistent synovitis where other pathologies have been ruled out needs to treated as if it is RA to try to prevent damage to joints. Identification of persistent synovitis and appropriate early management is more important than whether the disease satisfies classification criteria' referencing recommendations from the European League Against Rheumatism (EULAR).[15]

In 2010, the ACR and EULAR jointly published a Rheumatoid Arthritis Classification Criteria, which focussed on features at earlier stages of disease that are associated with persistent and/or erosive disease rather than defining the disease by its late stage features.[16] The classification criteria allocates scores to characteristics of: joint involvement; serology; acute-phase reactants; and duration of symptoms to produce a score between 0 and 10 inclusive, with those scoring 6 or greater and with obvious clinical synovitis being defined as having "definite RA" in the absence of an alternative diagnosis that better explains the synovitis.

Two classifications have dominated the measurement of improvement in RA symptoms: ACR responses[17] and EULAR responses.[18]

The initial ACR response was denoted as an ACR20 which required: a 20% improvement in tender joint counts; a 20% improvement in swollen joint counts; and a 20% improvement in at least three of

the following five 'core set items': Physician global assessment; Patient global assessment; patient pain; self-reported disability (using a validated instrument), and; erythrocyte sedimentation rate (ESR) / C-reactive protein.

ACR response has been widely adopted in randomised controlled trials (RCTs) although studies have shown that the value of the measure can vary between trials due to the timing of the response.[19] Since the inception of the ACR20, two further response criteria (ACR50 and ACR70) have become widely used. These are similar to ACR20 and differ only in the level of percentage improvements required to be classified as a responder.

In the UK, monitoring the progression of RA is often undertaken using the disease activity score of 28 joints (DAS28). This assesses 28 joints in terms of swelling (SW28) and of tenderness to the touch (TEN28) and also incorporates measures of the ESR and a subjective assessment (SA) on a scale of 0-100 made by the patient regarding disease activity in the previous week.

The equation for calculating DAS28 is as follows:[20]

$$DAS28 = 0.56* TEN28^{0.5} + 28* SW28^{0.5} + 0.70 * \ln (ESR) + 0.014 * SA$$

The DAS28 can be used to classify both the disease activity of the patient and the level of improvement estimated within the patient.

The EULAR response criteria use the individual change in DAS28 and the absolute DAS28 score to classify a EULAR response as good, moderate or none.[18] The EULAR response criteria and the ACR20 improvement criteria were found to have reasonable agreement in the same set of clinical trials, although van Gestel *et al.* state that the EULAR response criteria showed better construct and discriminant validity than ACR20.[21] EULAR response has been reported less frequently in RCTs than ACR responses,[22] although EULAR is much more closely aligned to the treatment continuation rules stipulated by NICE. These rules require either a moderate or good EULAR response or a DAS28 improvement of more than 1.2 to continue treatment, with the latter criterion applying to RTX. The relationship between change in DAS28 and the absolute DAS28 score and EULAR response is shown in

Table 1.

Table 1: Determining EULAR response based on DAS28[21]

	Improvement in DAS 28		
DAS28 at endpoint	>1.2	>0.6 and ≤1.2	≤0.6
≤ 3.2	Good	Moderate	None
>3.2 and ≤5.1	Moderate	Moderate	None
>5.1	Moderate	None	None

Patients with a DAS28 ≤3.2 are stated as having inactive disease, those with a DAS28 > 3.2 and ≤ 5.1 are stated as having moderate disease and >5.1 as having very active disease.[20]

A widely used measure of patient disability is the Health Assessment Questionnaire (HAQ). The HAQ score is a patient completed disability assessment which has established reliability and validity.[23] HAQ scores range from 0 to 3, with higher scores indicating greater disability, and is a discrete scale with step values of 0.125, resulting in the HAQ scale containing 25 points. The HAQ has been used in many published RCTs in RA.[22]

2.2 Critique of company's overview of current service provision

The company's overview of current service provision is concise but mostly appropriate and relevant to the decision problem in the final NICE scope. It is noted that the recommendations from TA375[24] have not been incorporated into the CS. The ERG provides a summary of current service provision below.

Clinical guidelines

For people with newly diagnosed RA, NICE CG79[13] recommends a combination of cDMARDs (including MTX and at least one other cDMARD plus short-term glucocorticoids) as first-line treatment, ideally beginning within 3 months of the onset of persistent symptoms. Where combination therapies are not appropriate, for example where there are comorbidities or pregnancy, cDMARD monotherapy is recommended. Where cDMARD monotherapy is used, emphasis should be made on increasing the dose quickly to obtain best disease control. For the purposes of this assessment, the term "intensive DMARDs" has been used to denote that this involves treatment with multiple cDMARDs simultaneously.

NICE guidance (Technology Appraisal (TA) 375)[24] recommends the use of ABA, ADA, CZP, ETA, GOL, IFX, and TOC in combination with MTX in people with RA after the failure to respond to intensive cDMARDs treatment and who have severe active RA (defined as a DAS28 score greater than 5.1). For people who meet this criteria but cannot take MTX because it is contraindicated or because of intolerance, TA375[24] recommends the following bDMARDs as monotherapy options: ADA; CZP; ETA; or TOC.

After the failure of the first TNF-inhibitor, TA195[25] recommends RTX in combination with MTX for the treatment of severe active RA. If RTX is contraindicated or withdrawn because of an adverse event (AE), TA195 recommends ABA, ADA, ETA, or IFX in combination with MTX. If MTX is contraindicated or withdrawn because of an AE, TA195 recommends ADA or ETA as monotherapy. TA247[26] recommends TOC as an alternative to TNF-inhibitors in the same circumstances as TA195, that is, after the failure of a TNF-inhibitor in patients with severe active RA, in combination with MTX when RTX is contraindicated or withdrawn and as monotherapy if MTX is contraindicated or withdrawn. In addition, TA247 recommends TOC in combination with MTX in patients in whom TNF-inhibitors and RTX have not worked.

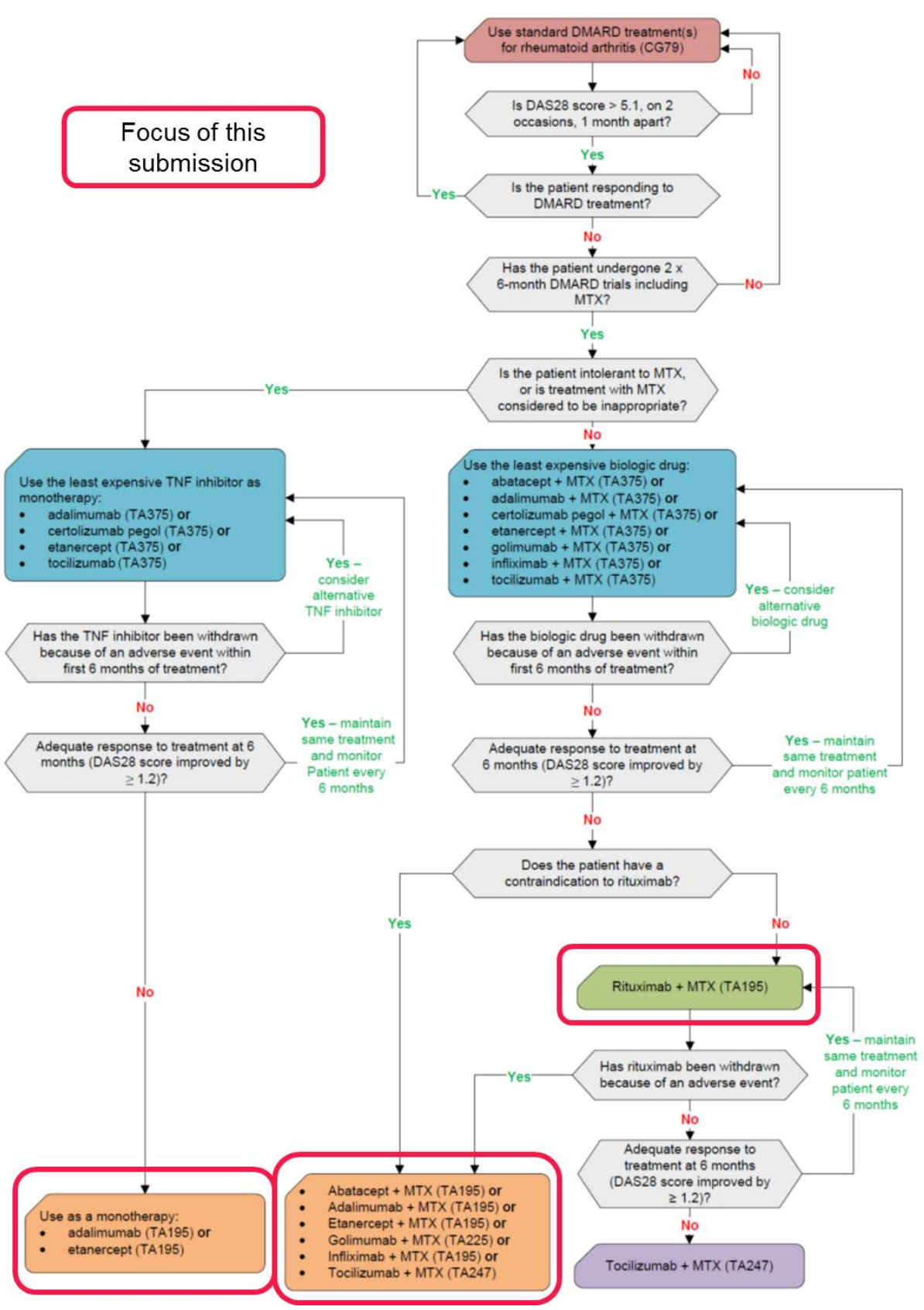
A simplified summary of NICE recommendations on bDMARDs is shown in

Figure 1. It defines the sequence of treatments that have received positive guidance for patients with a DAS28 score of >5.1. In summary, the typical route would be intensive cDMARDs followed by a bDMARD, followed by RTX plus MTX, then TOC before returning to cDMARDs.

NICE criteria for continuing treatment

NICE TA375[24] states that for patients to continue treatment with their first bDMARD treatment they must maintain at least a moderate EULAR response. TA195, which for all bDMARDs excluding RTX was updated in TA375[24], states that bDMARD treatment after the failure of a TNFi should be continued only if there is an adequate response (defined as an improvement in the DAS28 score of at least 1.2 points) at initiation of treatment and as long as this adequate response is maintained. If the criterion of having at least a moderate EULAR response at six months has not been met, then treatment should be stopped and the next intervention in the sequence initiated.

Figure 1: NICE recommended pathway for RA (adapted from Figure 2 in the CS)



3 CRITIQUE OF THE COMPANY'S DEFINITION OF THE DECISION PROBLEM

3.1 Population

The population identified in the NICE final scope[27] was 'Adults with moderate to severe, active rheumatoid arthritis whose disease has not responded adequately to a TNF inhibitor'. This definition may be not as was intended, as in TA375[24] patients with active RA were divided into two classifications: moderate to severe (defined as a DAS28 score >3.2 and ≤ 5.1) and severe (defined as a DAS28 >5.1). The ERG does not believe NICE intended to limit this STA to patients with a DAS28 score >3.2 and ≤ 5.1 and the company agreed. During the clarification process, the company clarified that the 'the patient characteristics in the model are based on the REALISTIC study, which enrolled a mixed population of patients with moderate (n=26, TNFi experienced) and severe RA (n=371, TNFi experienced)' (see clarification response: Question B1[28]). The company assumed that these patients could be analysed as one population rather than separating out the analysis into two populations: one for those with a DAS28 score >5.1 and one for those with a DAS28 score >3.2 and ≤ 5.1 which would be more consistent with the guidance in TA375.[24]

The company defined three subpopulations for the cost-effectiveness analysis based on contraindications to, or withdrawals due to AEs of, RTX and/or MTX:

A. Adults previously treated with other DMARDs including at least one TNFi
B. Adults for whom RTX is contraindicated or withdrawn
C. Adults for whom RTX therapy cannot be given because MTX is contraindicated or withdrawn

Further details on the populations assumed in the modelling are provided in Section 5.2.2.

3.2 Intervention

The intervention within the CS is in accordance with the final scope, namely 'Certolizumab pegol monotherapy or in combination with methotrexate'. CZP (CIMZIA®, UCB) is a PEGylated, fragment crystallisable-free, TNFi for the treatment of RA. It has a high affinity, neutralising potency, and specificity for human TNF-α, a key pro-inflammatory cytokine with a central role in inflammatory processes. CZP is the only PEGylated TNFi approved for the treatment of RA.[29] CZP is contraindicated in people with active tuberculosis or other severe infections, and in people with moderate or severe heart failure. The recommended starting dose of CZP for adult patients, known as the loading dose, is 400 mg (given as 2 subcutaneous (SC) injections of 200 mg each) at Weeks 0, 2 and 4. For RA, MTX should be continued during treatment with CZP where appropriate. After the loading dose, the recommended maintenance dose of CZP for adult patients with RA is 200 mg every 2 weeks (Q2W). Once clinical response is confirmed, an alternative maintenance dosing of 400 mg

every 4 weeks (Q4W) can be considered. MTX should be continued during treatment with CZP where appropriate. The price of a 200mg syringe prefilled with CZP is £375.50.[30] The company has agreed a PAS with the Department of Health where the first 12 weeks of therapy, consisting of 10 syringes pre-loaded with 200mg of CZP are provided to the NHS free of charge.

3.3 Comparators

Four comparators were defined in the final NICE scope, three of which were considered by the company. In population A, the intervention is added into a treatment sequence before RTX + MTX forming a comparison of an elongated sequence compared with a standard sequence. Other potential sequences, such as replacing RTX + MTX with CZP + MTX, or comparing the elongated sequence with an equally long sequence with RTX + MTX before CZP + MTX were not considered, and thus a fully incremental analysis of all sequences has not be undertaken within the CS.

For patients for whom RTX is contraindicated or has been withdrawn due to an AE, population B, only the first line of therapy is assumed to differ being either the intervention + MTX or one of the comparators named in the final scope (ABA, ADA, ETA, GOL, IFX, TOC) + MTX.

For patients for whom RTX therapy cannot be given because MTX is contraindicated or withdrawn, population C, only the first line of therapy is assumed to differ being either the intervention or one of the comparators named in the final scope (ADA, ETA, TOC).

The company did not incorporate the fourth comparator listed in the NICE final scope[27] which was best supportive care. The reason provided for the company was that this 'does not reflect the current NICE recommendations for TNFis (NICE Pathways for Drug Treatment for RA (26th March 2015); NICE commissioning algorithm (May 2013)). Furthermore, limited evidence supports the evaluation of CZP within this patient group.'

Further details on the comparators assumed in the company's health economic analysis are provided in Sections 5.2.3 and 5.2.8.

3.4 Outcomes

The majority of outcomes reported in the final NICE scope[27] have been included in the CS. Three defined outcome measures have, however, been excluded. These are: joint damage; radiological progression; and extra-articular manifestations of the disease.

4 CLINICAL EFFECTIVENESS

4.1 Critique of the methods of review(s)

This chapter presents the ERG critique and summary of the clinical effectiveness and safety evidence included in the CS.[31]

The clinical evidence presented in the CS comprised three main components:

1) A systematic review of clinical effectiveness evidence from CZP RCTs in patients with moderate to severe active RA with an inadequate response to TNFi treatment (TNFi-IR) (CS Section 4.7)

2) A review of observational (non-RCT) evidence for clinical effectiveness (CS Section 4.11)

3) A review of safety evidence from included CZP RCTs, a pooled analysis and three supporting safety studies (CS Section 4.12).

The searching and study selection processes for i) the review of observational (non-RCT) clinical effectiveness evidence and ii) the review of safety evidence were not clearly described in the CS. The supplementary non-RCT evidence was included to support the key clinical effectiveness and safety evidence from the included CZP RCTs.

4.1.1 Searches

The company conducted a systematic review of clinical effectiveness including trials of CZP and comparators (as specified in the final NICE scope[27]).

Embase and MEDLINE were cross-searched simultaneously via Embase.com. Although multi-file searching is generally not recommended for systematic reviews (due to the difficulties of optimising the search strategy to the specific features of each database) and the fact that the ERG has been unable to replicate the search results exactly as presented, the strategy appears to be well-structured and is likely to have been effective in identifying all relevant evidence.

A filter which was loosely based on those published by the Scottish Intercollegiate Guidelines Network (SIGN) was applied to restrict results to clinical trials. The ERG recognises SIGN as a respected source of expertise in information retrieval; however, it should be noted that their filters are not validated (tested for sensitivity and specificity against a gold standard). It should also be noted that, whilst simultaneously searching Embase and Medline (CS Section 8.2.1), the company appear to have relied mainly on a filter designed for searching Embase alone, and using Emtree subject headings, without also adding the equivalent MeSH headings to optimise the strategy for use in both databases.

The use of this filter also led the ERG to seek clarification as to how the company had searched for non-RCT evidence, given that observational studies had been identified. The company clarified (see clarification response, Question A2[28]) that some supplementary methods, including hand searching and horizon scanning, had been undertaken after the initial searches had been conducted.

The company also conducted searches of MEDLINE In Process and the CENTRAL Database of Clinical Trials (a section of the Cochrane Library). None of the other Cochrane Library databases (such as the Cochrane Database of Systematic Reviews) were searched.

Searches of conference proceedings (American College of Rheumatology [ACR] and the European League Against Rheumatism [EULAR]) were conducted; it might have also been useful to have searched registers of current clinical trials (e.g. ClinicalTrials.gov, clinicaltrialsregister.eu and the World Health Organisation's International Clinical Trials Registry Platform in order to identify additional completed or ongoing studies. It was not reported in the CS whether any forward citation tracking of key studies took place.

Searches were reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (http://prisma-statement.org/), although due to the multi-file searching issue detailed above, separate totals of records retrieved from Embase and Medline are not provided; instead, only a combined figure is reported.

Overall, the ERG is satisfied that the searches presented in the CS are likely to have identified all relevant published RCT evidence.

4.1.2 Inclusion criteria

The eligibility criteria applied in the clinical effectiveness systematic review (Table 2) were considered by the ERG to be reasonable and generally consistent with the decision problem specified in the final NICE scope.[27] In response to a request for clarification (see clarification response, Question A1[28]) the company confirmed that moderate to severe RA was defined as DAS28 > 3.2.

Interventions eligible for inclusion were bDMARDs (ABA, ADA, CZP, ETA, GOL, IFX, RTX, and TOC) administered as monotherapy or in combination with MTX in accordance with licensed indications.

The inclusion criteria for the clinical effectiveness systematic review prioritised the inclusion of disease activity outcomes (ACR response, EULAR response and DAS28). The decision problem

outlined in the final NICE scope[27] also listed physical function; joint damage; radiological progression; pain; mortality; fatigue; extra-articular manifestations of disease; and health-related quality of life (HRQoL) as relevant outcomes. Radiological progression and joint damage were not included as outcomes in the CS. However, data on inhibition of joint structural damage were reported in the published articles for both J-RAPID and HIKARI.[32, 33] Extra-articular manifestations of disease were not included in the CS.

The company's application of study design limits in the systematic review of clinical effectiveness (described in Appendix 8.3.1 of the CS) was considered valid by the ERG. Non-English studies were not eligible for the review. During clarification (see clarification response, Question A5[28]), the company stated that no relevant non-English language RCTs for CZP or comparators were identified.

The eligibility criteria for inclusion of evidence in the review of observational (non-RCT) clinical effectiveness evidence and the review of supplementary safety evidence were unclear.

Table 2: Eligibility criteria for systematic review of clinical effectiveness (adapted from CS Table 8)

Review component	Inclusion criteria	Exclusion criteria
Population	Adults (≥18 years) with moderate to severe active RA and TNFi-inadequate response	TNFi-naïve adult RA patients Paediatric RA patients
Intervention	•CZP monotherapy or + MTX •RTX + MTX •ABA, ADA, ETA, GOL, IFX and TOC + MTX •ADA monotherapy, ETA monotherapy or TOC monotherapy	Any other intervention
Comparators	• Any of above interventions • Best supportive care/cDMARDs/placebo (PBO)	Any other comparator
Outcomes	Disease characteristics: • ACR response rates (ACR20/50/70) • DAS28(ESR) remission (DAS28[ESR]<2.6) • DAS28(ESR) low disease activity (DAS28[ESR] <3.2) • EULAR response (none/good/moderate)	Any other outcomes
Study design	• RCTs (triple/double/single blind or open label) • Non-RCTs (CZP only): • Comparative cohort studies/longitudinal studies (prospective or retrospective) (CZP only) • Case-controlled studies (CZP only) •Comparative-cross-sectional studies (CZP only)	•Analysis of hospital records/database • Single arm studies (uncontrolled trials) • Case study • Case series • Case report • Systematic reviews
Language of publication	English language	Any other language

Study selection and data extraction

The study selection process was documented in a PRISMA flow diagram (CS Figure 3). Whilst the inclusion criteria for the clinical effectiveness systematic review indicated that specific types of non-RCT evidence would be eligible for the review, the number of included studies in the PRISMA diagram appears to represent the included RCTs only. Therefore, it was not clear to the ERG whether the PRISMA diagram represents RCTs only or also includes the study selection process for non-RCT evidence.

Six RCTs assessed CZP + MTX/cDMARDs or as monotherapy in patients with TNFi-IR (REALISTIC[34], DOSEFLEX[35], PREDICT[36], SWITCH[37], J-RAPID[32] and HIKARI[33]) (CS Table 9). The company noted that PREDICT did not strictly meet the inclusion criteria for the clinical effectiveness systematic review (as there was no non-CZP comparator arm). The company stated that there were no relevant ongoing trials for CZP (CS Section 4.14).

The ERG performed searches within the ClinicalTrials.gov trials register (March 2016) for completed interventional Phase 2/3/4 studies of CZP in RA. Twenty-nine studies were identified. The ERG also hand searched the Assessment Group report for the NICE multiple technology appraisal (MTA) of first-line use of biologics in RA for potentially relevant second-line RCTs. The European Medicines Agency (EMA[38]) Assessment Report for CZP was also checked by the ERG for any further relevant trials. In the cases of two CZP studies (NCT00753454 DOSEFLEX II; and RA0025 Kang *et al.* 2012[39]), the rationale for exclusion from the clinical effectiveness review was considered unclear by the ERG. DOSEFLEX II was referred to very briefly in the CS (page 50). Consequently, the ERG sought clarification from the company. In their clarification response (see clarification response, Question A8[28]), the company described DOSEFLEX II as a Phase IIIb, open-label CZP follow-up study to DOSEFLEX assessing safety and efficacy. Some supporting efficacy data from DOSEFLEX II were provided to the ERG. The company also responded that the CZP RCT reported by Kang *et al.* was not included in the CS because (of 121 patients included in the full analysis set) only 7 PBO + MTX (17.5%) patients and 11 CZP 200 mg Q2W + MTX (13.6%) patients were TNFi-experienced. However, the ERG noted that the company included in the CS two CZP RCTs that also had low numbers of TNFi-experienced patients: J-RAPID and HIKARI. Therefore, the study selection decision to exclude Kang *et al.*[39] could be viewed as inconsistent and omission from the clinical effectiveness review and network meta-analysis should have been further justified by the company. In the clarification responses, the company provided a pre-specified subgroup analysis from Kang *et al.*[39] of the primary outcome ACR20 at week 24. All data presented in the CS for HIKARI related to TNFi experienced patients receiving PBO or CZP in monotherapy only. Data for patients who received CZP in combination with non-MTX DMARDs were not presented in the CS. The company stated that the use of CZP in combination with non-MTX cDMARDs is not approved in the European Union.

No head-to-head RCTs were identified that directly compared the efficacy of CZP with other bDMARDs in TNFi-experienced patients.

The company clearly tabulated the nine RCTs included in the indirect treatment comparison in CS Appendix 8.12.1. Three trials were included for CZP + MTX (REALISTIC,[34] J-RAPID[32] and SWITCH[37]). Two RCTs were included for TOC + MTX (RADIATE,[40] Genovese *et al.*, 2014[41]); two studies for RTX + MTX (REFLEX[42], Combe *et al.*, 2012[43]); one for ETN + MTX (Combe *et al.*, 2012[43]); one RCT for ABA + MTX (ATTAIN[44]); and one trial for GOL + MTX (GO-AFTER[45]). It was stated in CS Section 4.10.3 that the Combe trial was only included in the sensitivity analysis due to small size and what the company described as implausible results. In their clarification response[28] (Question A26), the company expanded upon this rationale, explaining that the Combe trial had restricted inclusion criteria (TNFi-IR and RTX-IR) and that the quality assessment of this study suggested a high risk of bias. Since the Combe study was only available as a conference abstract, the ERG considered the inclusion of the Combe trial in the sensitivity analysis to be reasonable on the basis of limited methods and results details and unclear risk of bias. The company noted that, whilst trial data were available from REALISTIC and HIKARI for the use of CZP as monotherapy, indirect treatment comparisons were not performed since data were not available for any competitor treatment as monotherapy.

RCTs for CZP and comparators that were identified but excluded from the indirect treatment comparison were tabulated (CS Appendix 8.4.2) with reasons presented for their exclusion. These consisted of three CZP RCTs, namely PREDICT (Curtis *et al.* 2015[36]), HIKARI (Yamamoto *et al.* 2014b[33]), and DOSEFLEX (Furst *et al.* 2015[35]), and the following four trials relating to comparator bDMARDs: ASSURE (Weinblatt *et al.* 2006[34]), OPPOSITE (Furst *et al.* 2007[46]), SUNRISE (Mease *et al.* 2010[47]), and ROC (Gottenberg *et al.* 2015[48]).

The ERG requested clarification (see clarification response, Question A7[28]) from the company as to why the biologic comparator RCTs LITHE (TOC) and ORAL Standard (ADA), which studied mixed populations of biologic-naïve and biologic-experienced patients, were not included in the CS. The company stated that data from biologic-experienced patients were not available from these studies and so these were not included in the CS.

The clinical advisors to the ERG did not highlight any additional relevant RCTs that should have been included in the CS.

One relevant non-randomised study for CZP was included in the CS (see CS Section 4.11). The ARTIS observational study was based on the Swedish Rheumatology Quality Register database and provided supporting data on the clinical effectiveness of CZP.[49]

Safety data for CZP were also presented (CS Section 4.12) and were based on four of the included CZP RCTs (REALISTIC, DOSEFLEX, PREDICT and SWITCH). In response to a clarification request by the ERG (see clarification response, Question A16[28]) as to why the CZP RCTs J-RAPID and HIKARI were omitted from the safety review, the company provided additional data. A pooled analysis by Bykerk *et al.*[50] (assessing CZP safety from data from relevant RCTs and OLEs) was also included in the CS, alongside three additional studies describing the safety of CZP (Yun *et al.* 2016[51]; Simard *et al.* 2011[52], Curtis *et al.* 2015[36]). It appears that the pooled analysis by Bykerk *et al.*[50] utilised simple pooling for the combination of safety data from RCTs and OLEs (with it being stated in the study methods that patient data were pooled without the application of corrective factors).

It was not clear in the CS whether study selection and data extraction were undertaken by a single reviewer or involved two independent reviewers as per best practice.

Quality assessment

The quality of the six included CZP RCTs was assessed using appropriate and established criteria adapted from those recommended by the University of York Centre for Reviews and Dissemination (CRD). Results were summarised in CS Table 14. Randomisation was considered by the company to have been carried out appropriately in five of the six trials (with SWITCH marked unclear). Concealment of treatment allocation was judged to have been adequate in five trials (SWITCH being marked unclear). Treatment groups were considered to be similar at study outset in terms of (unstated) prognostic factors in all six trials. Care providers, participants and outcome assessors were judged to be blinded to treatment allocation in five trials (SWITCH marked as unclear). For all six trials, the company judged there to be no unexpected imbalances in drop-outs between groups. In terms of selective outcome reporting, for all six trials, there was no evidence to suggest study authors measured more outcomes than were reported in trial publications. Primary outcomes appeared consistent between trial publications and ClinicalTrials.gov records. For all six trials, it was stated by the company that there was use of intention to treat (ITT) analysis and appropriate methods to account for missing data. Quality assessment judgments were presented in full in CS Appendix 8.5. The grade of allocation concealment was recorded. The company also rated the CZP RCTs according to the Jadad scale,[53] with five trials awarded the highest score of 5 and a rating of 3 for SWITCH. The ERG considers the use of the Jadad scale to be of only limited value in usefully summarising the quality of trials.

The quality of the supporting observational study (ARTIS) was assessed using the Downs and Black checklist for non-randomised studies.[54] Quality findings were summarised in Section 4.11.5 of the CS and presented in full in Appendix 8.10. The company considered that the ARTIS publication included sufficient detail on: outcomes; patient characteristics; confounders; and findings. Whilst it was stated that ARTIS had good external validity (with the study cohort selected from a patient registry), internal validity was judged to be reduced due to bias in the study design from: lack of blinding; non-randomisation; and the lack of comparators. Statistical analyses and subgroup analyses were clearly described.

Quality assessment was performed for RCTs included in the indirect treatment comparison. Studies were appraised according to criteria based on the University of York Centre for Reviews and Dissemination guidance,[55] allocation concealment grade, and the Jadad scale.[53] Findings were reported in Appendix 8.12.1.3 of the CS.

4.1.3 Evidence synthesis

Study and patient characteristics for included CZP trials were clearly presented in a descriptive narrative summary alongside clinical and safety data. However, *p*-values were frequently not reported in the CS and therefore the ERG requested that these be provided by the company (see clarification response,[28] Questions A15 and A36).

Meta-analyses were performed for CZP used in combination with MTX and CZP as monotherapy. Classical meta-analyses were performed separately for: ACR20/50/70; EULAR response; and DAS28(ESR) remission at 3 months. Meta-analysis was not undertaken for outcomes at 6 months, since data were only available in J-RAPID for CZP + MTX and in HIKARI for CZP monotherapy. The ERG identified the availability of 24 weeks ACR and EULAR response data for SWITCH on ClinicalTrials.gov. However, since the PBO-controlled period only ran to week 12 of SWITCH (after which subjects received CZP in an OLE), the ERG considered it appropriate that 24-week data for SWITCH not be included in meta-analyses.

Both fixed effects (Mantel-Haenszel) and random effects (DerSimonian and Laird) models were used. Heterogeneity between trials was investigated using I^2 values. It is generally recommended that at least five studies should be available for a classical meta-analysis, whereas the analyses included, at most, only three studies; the DerSimonian and Laird estimate of the between-study standard deviation over-estimates the population between-study standard deviation on average, and when there are few studies the bias can be substantial.

A Binomial likelihood assuming fixed effect and random effects models for parameters were used to model the separate ACR and EULAR categories. However, reference prior distributions were used for the between-study standard deviations despite there being an insufficient number of studies to allow Bayesian updating; the prior distributions were inappropriately informative leading to implausible posterior results on which to base inferences. A fixed effect model was used to model the ACR and EULAR data allowing for their ordered categories with no allowance for heterogeneity. In general, heterogeneity would be expected and the ERG would have preferred to see a random effects model incorporating plausible weakly informative prior information for the between-study standard deviation. In the presence of heterogeneity, the treatment effects from a random effects model do not represent the treatment effects in any specific patient population and it is recommended that inferences are made based on the predictive distribution of the treatment effects in a future study.[56] The results from a fixed effect model answers the question, "Did the treatments have an effect in the studies included in the analysis?" and, relative to a random effects model, under-estimates uncertainty when heterogeneity is expected.

4.2 Critique of trials of the technology of interest, their analysis and interpretation

4.2.1 Study design characteristics of included CZP RCTs

The study design characteristics of the included CZP trials are presented in Table 3. All six trials were funded by UCB Pharma, with two studies (J-RAPID and HIKARI) performed solely in Japan. All trials were performed in DMARD-experienced RA patients and typically included a mix of both TNFi-naïve and TNFi-experienced patients (with the exception of SWITCH being performed in TNFi-experienced patients only). Five of the RCTs were PBO-controlled, with the exception being PREDICT, which did not include a non-CZP comparator arm. CZP was administered in combination with background cDMARDs or monotherapy as detailed in study descriptions. REALISTIC, PREDICT and HIKARI provided data for the use of CZP as monotherapy. DOSEFLEX was the only included CZP RCT to have an open-label run-in phase in which all patients received CZP. Only ACR20 responding patients were randomised into the PBO-controlled double-blind period of DOSEFLEX.

The company clarified (see clarification response, Question A17[28]) that none of the included UCB Pharma-initiated CZP RCTs included dose adjustments of CZP. The company also confirmed (see clarification response, Question A17[28]) that only J-RAPID and HIKARI permitted poorly responding patients to enter early escape (ACR20 non-responders at Weeks 12 or 14 were withdrawn at Week 16 and were eligible to enrol in open-label extension (OLE) studies receiving CZP 200 mg Q2W). Patients who received rescue medication were considered non-responders from that point forwards.

In their response to the ERG's request for clarification (see clarification response, Question A10[28]), the company stated that none of the six included CZP RCTs included subjects who were recruited from centres in the UK.

4.2.1.1 REALISTIC

REALISTIC (RA Evaluation in Subjects Receiving TNF Inhibitor CZP) was a Phase IIIb RCT intended to reflect clinical practice in which the efficacy and safety of CZP as monotherapy or in combination with current cDMARD treatments was studied in a moderate to severe DMARD-experienced RA population. The trial was conducted in the USA, Canada and Europe. Of 1063 patients (CZP=851, PBO=212), 37.6% had previous TNFi use. Patients were randomised 4:1 to CZP (400 mg SC at weeks 0, 2 and 4 then 200 mg Q2W) or PBO respectively. A 12-week double-blind phase was followed by an OLE in which patients received CZP 200 mg Q2W. The REALISTIC trial was the largest of the included studies in terms of participant numbers. REALISTIC excluded patients with: Prior receipt of either >2 TNFis or RTX and/or ABA or discontinuation of a biologic therapy for RA due to severe hypersensitivity or anaphylactic reactions.

4.2.1.2 DOSEFLEX

The purpose of the DOSEFLEX (dosing flexibility) trial was to assess the clinical efficacy and safety of two CZP maintenance dosing regimens in combination with MTX in ACR20 responders after an open-label run-in period. The trial was conducted at 63 centres across the USA, Canada and France. DMARD-experienced subjects with moderate to severe RA receiving stable doses of MTX entered an open-label run-in phase involving administration of CZP 400 mg at weeks 0, 2 and 4 and at a dose of 200 mg Q2W to week 16. At week 18, ACR20 responders at week 16 (209 of 333 patients) were randomised 1:1:1 into the PBO-controlled double-blind phase to receive either CZP 400 mg Q4W + MTX; CZP 200 mg Q2W + MTX; or PBO + MTX up to week 34. ACR20 non-responders were withdrawn from the trial at week 16.

4.2.1.3 PREDICT

The aim of the PREDICT trial (Patient/Physician Reported Efficacy Determination in Clinical Practice Trial) was to compare two tools in the assessment of CZP treatment response at week 12 and in the prediction of treatment response at 52 weeks in DMARD-experienced patients with moderate to severe RA. The trial was based at 110 centres in the USA. Patients were randomised to RAPID3 (subject-based tool) or Clinical Disease Activity Index (CDAI) (investigator-based tool) to determine patient response. It was noted in the CS that this trial did not contain a non-CZP comparator arm and therefore does not permit the comparison of CZP with any of the relevant comparators detailed in the decision problem in the final NICE scope.[27] All patients received open-label CZP 400 mg at weeks 0, 2 and 4 then CZP 200 mg Q2W from week 6 through week 52.

4.2.1.4 SWITCH

The SWITCH trial was undertaken at 12 centres in the USA to investigate the efficacy and safety of CZP in RA patients on stable concomitant cDMARDs who had discontinued prior TNFi treatment for secondary loss of efficacy or lack of tolerability. Thirty-four patients had experienced previous loss of efficacy and 3 patients were intolerant (migraine n=1, injection site reaction n=1, infusion reaction and injection site reaction to another TNFi n=1). Patients were allowed to have failed more than one TNFi. Patients were required to have undergone a washout from the previous TNFi of ≥4 weeks prior to their baseline visit. Patients were randomised 2:1 into a 12-week double-blind phase to receive CZP (400 mg SC at weeks 0, 2 and 4 then 200 mg Q2W) (n=27) or PBO (n=10) followed by a 12 week open-label phase. Schiff *et al.* (2014[37]) stated that, whilst it was originally planned that 102 patients would be randomised, study enrolment was halted by the sponsor-investigator on the justification that an interim analysis had shown that the primary hypothesis had been met. SWITCH was an investigator-initiated study, and UCB Pharma clarified (see clarification response, Question A12[28]) that they did not hold the trial data.

4.2.1.5 J-RAPID

The Japan RAPID (J-RAPID) 24-week trial was performed at 67 centres across Japan to determine the efficacy, pharmacokinetics and safety of CZP in combination with MTX in Japanese DMARD-experienced RA patients with an inadequate response to MTX. 316 patients were randomised 1:1:1:1 to CZP SC 100, 200 or 400 mg (induction dose 200 mg or 400 mg at weeks 0, 2 and 4) + MTX or PBO + MTX Q2W. ACR20 non-responders at weeks 12 and 14 were withdrawn from the trial at week 16 and could enter an OLE study.

4.2.1.6 HIKARI

The objectives of the 24-week HIKARI trial (conducted in Japan) were to confirm the superiority in ACR20 efficacy and to assess the pharmacokinetics and safety of CZP versus PBO without co-administration of MTX (i.e. CZP monotherapy) in DMARD-experienced RA patients who could not receive MTX (because of inadequate efficacy, safety issues or previous discontinuation for safety reasons). Patients were randomised for treatment with CZP or PBO in monotherapy or in combination with non-MTX DMARDs. 230 subjects were randomised 1:1 to CZP 200 mg Q2W or PBO Q2W. ACR20 non-responders at weeks 12 and 14 were withdrawn from the study at week 16 and could enter an OLE study.

Table 3: Study characteristics of included CZP RCTs (adapted from CS Tables 9, 11 and 12)

Trial name / Author, year (NCT/sponsor number) **Primary publication details / Funding source**	**Trial design**	**Treatment arms included in CS (number of patients randomised per treatment arm)**	**Study duration**	**Key subject inclusion criteria**	**Key subject exclusion criteria**	**Early rescue/escape plan reported?**	**Geographical location**
REALISTIC (NCT00717236) Weinblatt *et al.* 2012[34] / UCB Pharma	RCT Randomised, double-blind, PBO-controlled, parallel group, stratification based on concomitant use of MTX; prior TNFi use; disease duration <2 or ≥ 2 yrs Phase IIIb	CZP 400 mg (SC) at weeks 0, 2 and 4 then CZP 200 mg Q2W +/-MTX/cDMARDs (n=851, 320 TNF-experienced) PBO (SC) +/-MTX/cDMARDs (212, 80 TNF-experienced) Analgesics, oral corticosteroids (≤ 10 mg/day prednisone equivalent) and NSAIDs/COX-2 inhibitors permitted if doses stable within 24 h, 7 days and 14 days respectively of baseline	28 weeks (12 week double-blind phase followed by open-label CZP 200 mg Q2W)	• Aged ≥18 yrs • Adult onset RA (ACR definition) for ≥3 months • Active moderate to severe RA defined as: - ≥5 tender joints (28 joint count) AND - ≥4 swollen joints (28 joint count) AND - At least one of the following two criteria: - CRP >10 mg/L - ≥28 mm/h ESR (Westergren) • Failure to respond (lack of efficacy or intolerance) to ≥1 synthetic DMARD • Discontinuation of all ineligible DMARD therapy ≥28 days / 5 half-lives prior to first dose of study drug (whichever longer)	• Any other inflammatory arthritis • History of chronic, serious or life-threatening infection; any current infection; active or history of active tuberculosis, • Receipt of biologic therapy for RA within 2 months (1 month for ETA or anakinra) prior to baseline • Prior receipt of either >2 TNFis or RTX and/or ABA or discontinuation of a biologic therapy for RA due to severe hypersensitivity or anaphylactic reactions	No	230 centres in USA and Canada (75%) and Europe (25%) [France, Germany, Italy, Netherlands and Spain)
DOSEFLEX (NCT00580840) Furst *et al.* 2015[35] / UCB Pharma	RCT Randomised, double-blind, PBO-controlled,	Open-label run in phase: CZP 400 mg (SC) at weeks 0, 2 and 4	34 weeks (Open-label CZP run-in phase to week 16	• Aged ≥18 yrs • Diagnosis of adult-onset moderate to severe RA, having lasted 6 months-15 years (ACR 1987	• Other inflammatory arthritis (or secondary, non-inflammatory arthritis) • Use of prohibited medication (analgesics,	No	63 centres in USA, Canada and France

Trial name / Author, year (NCT/sponsor number) Primary publication details / Funding source	Trial design	Treatment arms included in CS (number of patients randomised per treatment arm)	Study duration	Key subject inclusion criteria	Key subject exclusion criteria	Early rescue/escape plan reported?	Geographical location
	parallel group, open-label run-in Phase IIIb	CZP 200 mg Q2W + MTX (70, 43 TNF-experienced) CZP 400 mg Q4W + MTX (70, 39 TNF-experienced) PBO + MTX (69, 29 TNF-experienced) Concomitant treatment with analgesics, NSAIDs/COX-2 inhibitors and corticosteroids (≤ 10 mg/day or equivalent) permitted.	then double-blind, PBO-controlled randomised period from weeks 16 to 34)	criteria) • RF positive and/or anti-CCP positive • Treated with MTX (10 to 25 mg/week, +/- folic acid) for ≥ 3 months prior to the baseline visit • Active RA defined as: - ≥6 tender joints (28 joint count) AND - ≥4 swollen joints (28 joint count) AND - At least one of the following: - CRP >10 mg/L - ≥28 mm/h ESR (Westergren)	oral corticosteroids, IA hyaluronic acid, specific DMARDs) • Failure to respond to previous treatment with TNFis (primary failures) • Chronic infection, recent serious or life-threatening infection, or current sign of infection; high risk of infection; history of active or latent tuberculosis • History of lymphoprolif-erative disorder		

Trial name / Author, year (NCT/sponsor number) Primary publication details / Funding source	Trial design	Treatment arms included in CS (number of patients randomised per treatment arm)	Study duration	Key subject inclusion criteria	Key subject exclusion criteria	Early rescue/escape plan reported?	Geographical location
PREDICT (NCT01255761) Curtis *et al.* 2015[36] / UCB Pharma	RCT (Randomised, double-blind, parallel group) Phase IV	CZP 400 mg (SC) at weeks 0, 2 and 4 then CZP 200 mg Q2W +/- MTX/cDMARDs from week 6 through week 52 All patients received open-label CZP and were randomised to either RAPID3 (368, 194 TNF-experienced) or CDAI (365, 213 TNF-experienced) assessment tools. Concomitant treatment with stable doses of cDMARDs permitted.	52 weeks	• Aged ≥18 years • A diagnosis of adult-onset RA of ≥3 months' duration at baseline (defined as 1987 ACR classification criteria): • ≥4 tender joints (28 joint count) • ≥4 swollen joints (28 joint count) • Unsatisfactory response or intolerance to ≥1 DMARD • Treatment with ≤2 TNFis prior to enrolment	• A diagnosis of any other inflammatory arthritis • A diagnosis of secondary, non-inflammatory type of arthritis, symptomatic enough to interfere with evaluation of the effect of CZP • A diagnosis of fibromyalgia with sufficient symptoms requiring treatment	No	110 centres in US
SWITCH (NCT01147341) Schiff *et al.* 2014[37] / UCB Pharma	RCT (Randomised, double-blind, PBO-controlled, parallel group) Phase IV	CZP 400 mg (SC) at weeks 0, 2 and 4 then CZP 200 mg Q2W + cDMARDs (27, all TNF-experienced) PBO + cDMARDs (10, all TNF-experienced) Receipt of MTX/other DMARDs was permitted at same dose prior to study entry. Oral corticosteroids (≤ 10 mg/day prednisone	24 weeks (12 week double-blind phase followed by 12 week open-label phase)	• Aged 18 to 75 years • RA (1987 ACR criteria) of >6 months duration, functional Class 1 – 3. • Active RA defined as: - ≥6 tender joints (28 joint count) AND - CRP >10 mg/L - CDAI ≥12 at screening • Previous secondary inadequate response or intolerant of a TNFi	• B cell depleting agent taken within 6 months prior to enrolment • No significant response to prior TNFi	No	12 centres in US

Trial name / Author, year (NCT/sponsor number) Primary publication details / Funding source	Trial design	Treatment arms included in CS (number of patients randomised per treatment arm)	Study duration	Key subject inclusion criteria	Key subject exclusion criteria	Early rescue/escape plan reported?	Geographical location
		equivalent) permitted with stable dose within 1 month of baseline and during study.		other than CZP • Oral MTX (≥ 10 mg/week) or other cDMARD if MTX intolerant continuously for ≥3 months before first study dose			
J-RAPID (NCT00791999) Yamamoto *et al.* 2014a[32] / Sponsor Otsuka Pharmaceutical Co., Ltd and collaborator UCB Japan Co. Ltd.	RCT (Randomised, double-blind, placebo-controlled, parallel group) Phase II/III	CZP 200 mg Q2W + MTX (82, 11 TNF-experienced) PBO + MTX (77, 15 TNF-experienced) MTX dose set to 6-8 mg/wk in line with licensed dose in Japan at time of trial	24 weeks	• Aged 20-75 yrs • RA (ACR 1987 criteria) of 0.5−15 years' duration • Treated with MTX for ≥6 months with fixed dose for prior ≥2 months at 6-8 mg/week • Active RA defined as: - ≥9 tender joints (68 joint count) AND - ≥9 swollen joints (66 joint count) - At least one of the following: - ESR ≥30 mm/hour - CRP ≥1.5 mg/dL	• Any biologic therapy for RA within the 6 months before study (3 months for ETA) • Any investigational drug in preceding 3 months • Previous severe hypersensitivity or anaphylactic reaction following TNFi treatment • Previous treatment with ≥2 TNFis; or previously failed to respond to TNFi therapy	Yes	67 centres in Japan
HIKARI (NCT00791921) Yamamoto *et al.* 2014b[33] / Sponsor Otsuka	RCT (Randomised, double-blind, placebo-controlled, parallel group)	CZP 400 mg at weeks 0, 2, and 4 then CZP 200 mg Q2W +/- non-MTX cDMARDs (116) PBO +/- non-MTX	24 weeks	• Aged 20-74 yrs • RA (ACR 1987 criteria) of 0.5-15 years' duration • Failed treatment with or resistant to ≥1 prior DMARD (including	• Non-RA inflammatory arthritis • Any biologic for RA in the 6 months preceding the study (3 months for ETA) • Any investigational drug	Yes	Japan

Trial name / Author, year (NCT/sponsor number) Primary publication details / Funding source	Trial design	Treatment arms included in CS (number of patients randomised per treatment arm)	Study duration	Key subject inclusion criteria	Key subject exclusion criteria	Early rescue/escape plan reported?	Geographical location
Pharmaceutical Co., Ltd and collaborator UCB Japan Co. Ltd.	Phase III	cDMARDs (114) Concomitant non-MTX DMARDs, NSAIDs/COS-2 inhibitors, corticosteroids (to 10 mg/day prednisone equivalent) were permitted at stable doses.		MTX), or unable to receive MTX due to safety concerns • Non-MTX DMARDs permitted if fixed dose for prior ≥2 months • Active RA defined as: - ≥6 tender joints (68 joint count) AND - ≥6 swollen joints (66 joint count) - At least one of the following: - ESR ≥28 mm/hour - CRP ≥2.0 mg/dL	in the preceding 3 months • Previous treatment with ≥2 TNFis, or previously failed to respond to TNFi therapy • Previous severe hypersensitivity or anaphylactic reaction following bDMARDs		

4.2.2 Participants' baseline characteristics

Baseline patient characteristics for the six included CZP RCTs are summarised in Table 4. In the REALISTIC trial, randomisation was stratified according to baseline MTX use, prior TNFi use, and disease duration (<2 years vs. ≥2 years). The baseline characteristics of patients entering the open-label and randomised phases of the DOSEFLEX trial were broadly comparable. Where reported in the included trials, study participants were [redacted]Mean disease duration ranged from[redacted]). Mean HAQ Disability Index (HAQ-DI) scores ranged from[redacted]Mean DAS-28 (ESR) scores were reasonably[redacted] In response to a request for clarification from the ERG (see clarification response, Question C2[28]), the company provided the ERG with revised figures for the proportions of patients who were Rheumatoid Factor (RF)-positive at baseline. In the majority of trials, [redacted] The previous DMARD treatment history among the overall mixed study populations is presented in Table 5. The mean dose of MTX at baseline was lower in J-RAPID, with the dose set to 6-8 mg/week in line with the licensed dose in Japan at the time of study. In their responses to the ERG's request for clarification (see clarification response, Question A9[28]), the company provided more specific details of prior biologic exposure, which are reproduced in Table 6 and Table 7. [redacted]

Table 4: Baseline patient characteristics of included CZP RCTs (adapted from CS Table 13)

Trial name	Treatment arm (n)	n [% TNFi experienced]	Mean age (SD), years	Female, n (%)	Mean disease duration, years (SD)	HAQ-DI mean (SD)	DAS28(ESR), mean (SD)	RF-positive (≥14 IU/mL), n (%)**
REALISTIC	CZP 200 mg Q2W +/- MTX/cDMARDs	TNFi experienced (n=320) [37.6%]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	PBO +/- MTX/cDMARDs	TNFi experienced (n=80) [37.7%]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
DOSEFLEX	CZP 200 mg Q2W + MTX	TNFi experienced (n=43) [61.4%]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	CZP 400 mg Q4W + MTX	TNFi experienced (n=39) [55.7%]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	PBO + MTX*	TNFi experienced (n=29) [42.0%]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
PREDICT	CZP 200 mg Q2W +/- MTX/cDMARDs	TNFi experienced (n=407) [55.5%]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
SWITCH	CZP 200 mg Q2W + cDMARDs	TNFi experienced (n=27) [100.0%]	56.1	NR	12.0	1.5	5.5	NR
	PBO + cDMARDs	TNFi experienced (n=10) [100.0%]	59.0	NR	14.0	1.1	5.4	NR

Trial name	Treatment arm (n)	n [% TNFi experienced]	Mean age (SD), years	Female, n (%)	Mean disease duration, years (SD)	HAQ-DI mean (SD)	DAS28(ESR), mean (SD)	RF-positive (≥14 IU/mL), n (%)**
J-RAPID	CZP 200 mg Q2W + MTX	Overall patients (n=82[redacted]	50.6 (11.4)	69 (84.1)	5.6 (4.2)	1.1 (0.7)	6.2 (0.8)	71 (86.6)
	PBO + MTX	Overall patients (n=77) [redacted]	51.9 (11.1)	66 (85.7)	5.8 (4.1)	1.2 (0.7)	6.5 (0.9)	66 (85.7)
HIKARI	CZP 200 mg Q2W -/+ non-MTX cDMARDs	Overall patients (n=116) [6.9%]	56.0 (10.2)	83.7 (71.6)	5.4 (4.0)	1.05 (0.7)	6.1 (0.9)	99 (85.3)
	PBO -/+ non-MTX cDMARDs	Overall patients (n=114) [14.0%]	55.4 (9.8)	88 (77.2)	5.8 (4.3)	1.21 (0.7)	6.3 (1.0)	102 (89.5)

* Baseline characteristics for TNFi-experienced PBO group patients in DOSEFLEX were provided by the manufacturer in response to a request by the ERG and have been included by the ERG in the table above. Data represented patients who completed run-in phase and were randomised into the double-blind phase

** RF-positive data were revised by the company in a request for clarification by the ERG and are included in the table above

Table 5: Previous DMARD treatment history among overall mixed study populations

Trial name	Treatment arm (n)	Prior DMARD treatment history (total population)
REALISTIC	CZP 200mg Q2W +/- MTX/cDMARDs	MTX use at baseline, n (%) = 589 (69.2); mean (SD) dose, mg/week = 17.2 (5.7) Concomitant DMARD use, n (%) = 697 (81.9) Previous TNFi use, n (%) = 320 (37.6)
	PBO +/- MTX/cDMARDs	MTX use at baseline, n (%) = 143 (67.5); mean (SD) dose, mg/week = 16.6 (5.3) Concomitant DMARD use, n (%) = 165 (77.8) Previous TNFi use, n (%) = 80 (37.7)
DOSEFLEX	CZP 200mg Q2W + MTX	Double-blind phase: Concomitant MTX dose, mg/week (SD) = 17.5 (4.3) Previous TNFi use, n (%) = 43/70 (61.4)
	CZP 400mg Q4W + MTX	Concomitant MTX dose, mg/week (SD) = 18.0 (5.2) Previous TNFi use, n (%) = 39/70 (55.7)
	PBO + MTX	Concomitant MTX dose, mg/week (SD) = 16.6 (4.8) Previous TNFi use, n (%) = 29/69 (42.0)
PREDICT	CZP 200mg Q2W +/- MTX/cDMARDs	Previous TNFi use, n (%) = 407/733 (55.5)
SWITCH	CZP 200mg Q2W + cDMARDs	Average no. of TNFi failed = 1.4 Prior ETN treatment (n) = 11; prior IFX treatment (n) = 12; prior ADA treatment (n) = 13; prior GOL treatment (n) = 1 Reason to stop prior TNFi (loss of efficacy/incomplete response/other) = 33/3/2 MTX dose (mean mg/week) = 16.4
	PBO + cDMARDs	Average no. of TNFi failed = 1.33 Prior ETN treatment (n) = 4; prior IFX treatment (n) =7; prior ADA treatment (n) =2; prior GOL treatment (n) =1 Reason to stop prior TNFi (loss of efficacy/incomplete response/other) = 11/0/2 MTX dose (mean mg/week) = 16.1
J-RAPID	CZP 200mg Q2W + MTX	Mean number of prior DMARDs (SD), incl MTX = 1.7 (0.8); mean MTX dose (SD), mg/week = 7.6 (0.8) Baseline CS use, n (%) = 56/82 (68.3) Prior TNFi use, n (%) = 11/82 (13.4)

Trial name	Treatment arm (n)	Prior DMARD treatment history (total population)
	PBO + MTX	Mean number of prior DMARDs (SD), incl MTX = 1.8 (0.9); mean MTX dose (SD), mg/week = 7.4 (0.9) Baseline CS use, n (%) = 46/77 (59.7) Prior TNFi use, n (%) = 15/77 (19.5)
HIKARI	CZP 200mg Q2W -/+ non-MTX cDMARDs	Mean number of prior DMARDs (SD), incl MTX = 1.9 (1.0) Baseline CS use, n (%) = 77/116 (66.4) Prior TNFi use, n (%) = 8/116 (6.9)
	PBO -/+ non-MTX cDMARDs	Mean number of prior DMARDs (SD), incl MTX = 1.8 (0.9) Baseline CS use, n (%) = 81/114 (71.1) Prior TNFi use, n (%) = 16/114 (14.0)

Table 6: Baseline characteristics: summary of prior TNFi exposure in CZP studies‡ (reproduced from company's clarification response to ERG)

	REALISTIC*		**DOSEFLEX‡**			**PREDICT***	**J-RAPID***		**HIKARI***	
	CZP 200 mg Q2W (n=851)	PBO (n=212)	CZP 200 mg Q2W (n=70)	CZP 400 mg Q4W (n=70)	PBO (n=69)	CZP 200 mg Q2W (n=733)	CZP 200 mg Q2W (n=82)	PBO (n=77)	CZP 200 mg Q2W (n=116)	PBO (n=114)
Proportion of patients with prior TNFi exposure, n (%)										
Adalimumab	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Etanercept	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Golimumab	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Infliximab	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

*****Full analysis set; ‡Enrolled set

‡Data presented in this table refers to: past TNFis (REALISTIC), prior TNFis (DOSEFLEX, PREDICT) and pre-treatment (J-RAPID and HIKARI). Past TNFi refers as any TNFi taken in the past in REALISTIC. Prior TNFi refers to any TNFi taken in the past in DOSEFLEX and PREDICT. Pre-treatment refers to any anti-rheumatic treatments taken in the past in J-RAPID and HIKARI.

Table 7: Baseline characteristics: summary of duration of exposure for prior TNFis in CZP studies‡ (reproduced from company's clarification response to ERG)

	PREDICT*	J-RAPID*		HIKARI*	
	CZP 200 mg Q2W (n=733)	CZP 200 mg Q2W (n=82)	PBO (n=77)	CZP 200 mg Q2W (n=116)	PBO (n=114)
Duration of exposure, mean days (SD)					
Adalimumab	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Etanercept	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Golimumab	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Infliximab	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

*****Full analysis set; ‡Enrolled set

‡Data presented in this table refers to: past TNFis (REALISTIC), prior TNFis (DOSEFLEX/PREDICT) and pre-treatment (J-RAPID and HIKARI). Past TNFi refers as any TNFi taken in the past in REALISTIC. Prior TNFi refers to any TNFi taken in the past in DOSEFLEX and PREDICT. Pre-treatment refers to any anti-rheumatic treatments taken in the past in J-RAPID and HIKARI.

4.2.3 Participant flow and numbers

Five CZP RCTs included mixed populations of TNFi-naïve and TNF-experienced patients (with the exception being SWITCH, which included solely TNFi-experienced subjects). REALISTIC was the largest study, with 851 (320 TNFi-experienced) and 212 (80 TNFi-experienced) patients randomised to CZP and PBO, respectively.

In the REALISTIC trial, a high proportion (≥ 85%) of patients in both CZP and PBO treatment arms completed the double-blind controlled phase to 12 weeks, with occurrence of AEs and lack of efficacy being the most common specified reasons for discontinuation. The majority (≥ 80%) of randomised patients also completed the open-label period. The majority of patients (≥ 78%) randomised at week 18 of DOSEFLEX also completed the double-blind phase to week 34 of the study. AEs were the most common reason for withdrawal in the CZP treatment arms, whilst (as might be anticipated) loss of efficacy contributed to most drop-outs in the PBO arm. It is worth noting that DOSEFLEX included an open-label CZP run-in phase that resulted in a considerable number of patients (93/333 [27.9%]) who entered the run-in phase dropping out due to lack of efficacy and not being randomised into the double-blind period. It is also interesting that, whilst a large proportion of patients completed to week 12 of PREDICT (≥ 80% in both CZP groups), only approximately 50% remained at week 52, with lack of efficacy being cited as the most frequent reason for CZP discontinuation. Over 90% of randomised patients in the SWITCH trial completed 12 weeks of the study. Whilst the completion rate of randomised patients to 24 weeks was good (≥ 70%) in the CZP arms of J-RAPID and HIKARI, only 35.5% and 15.8% of PBO group patients completed (with most withdrawing at week 16 due to lack of efficacy).

In most CZP RCTs (REALISTIC, PREDICT, DOSEFLEX, J-RAPID and HIKARI), missing data were imputed using the last observation carried forward (LOCF) or non-responder imputation (NRI) methods (for continuous and categorical outcome measures respectively). Data imputation methods used in SWITCH were not reported by Schiff *et al.*[37] and no further details could be identified in the trial record on ClinicalTrials.gov.

Table 8: Participant flow in included CZP trials

Trial	Arms	Randomised, n	Completed double-blind phase, n (%)	Discontinued prior to end of double-blind phase, n (%)	Analysis sets, n (%)	Entered OLE phase	Completed OLE, n (%)	Discontinued prior to end of OLE, n	Analysis sets, n (%)
REALISTIC	CZP 200 mg Q2W +/ DMARDs	851 (320 TNFi-experienced)	To 12 weeks 771 (90.6) (286 TNFi-experienced)	80 (9.4) AE = 33 (3.9) Lack of efficacy = 6 (0.7) Loss of efficacy = 3 (0.4) LtFU = 3 (0.4) WoC = 10 (1.2) Other unspecified = 25 (2.9)	Safety set: 846 (99.4) Full analysis set: 851 (100) Per protocol set: 629 (73.9)	771 (90.6) (286 TNFi-experienced)	Until ≥ week 28 720 (84.6) (286 TNFi-experienced)	Prior to week 28 51 AE = 15 Lack of efficacy = 10 Loss of efficacy = 1 LtFU = 13 WoC = 7 Other unspecified = 5	Open-label set: 770 *
	PBO +/ DMARDs	212 (80 TNFi-experienced)	To 12 weeks 184 (86.8) (80 TNFi-experienced)	28 (13.2) AE = 6 (2.8) Lack of efficacy = 6 (2.8) Loss of efficacy = 0 LtFU = 5 (2.4) WoC = 2 (0.9) Other unspecified = 9 (4.2)	Safety set: 209 (98.6) Full analysis set: 212 (100) Per protocol set: 166 (78.3)	184 (86.8) (80 TNFi-experienced)	Until ≥ week 28 175 (82.6) (TNFi-experienced NR)	Prior to week 28 9 AE = 2 Lack of efficacy = 5 Loss of efficacy = 0 LtFU = 2 WoC = 0 Other = 0	Open-label set: 184
DOSEFLEX **	CZP 200 mg Q2W + MTX	Randomised at week 18 70 (43 TNFi-experienced)	To 34 weeks 61 (87.1) (43 TNFi-experienced)	Prior to week 34 9 (12.9) AE = 4 Loss of efficacy = 2 Withdrawal by subject = 2 Forbidden medication = 1	70	NR	NR	NR	NR
	CZP 400 mg Q4W + MTX	Randomised at week 18 70 ǂ (39 TNFi-experienced)	To 34 weeks 63 (90.0) (39 TNFi-experienced)	Prior to week 34 7 (10.0) AE = 3 Loss of efficacy = 1 LtFU = 1 Withdrawal by subject = 1 Discontinued MTX = 1	69	NR	NR	NR	NR
	PBO Q2W + MTX	Randomised at week 18 69 (29 TNFi-experienced)	To 34 weeks 54 (78.3) (28 TNFi-experienced)	Prior to week 34 15 (21.7) AE = 1 Loss of efficacy = 10 LtFU = 2 Withdrawal by Subject = 2	69	NR	NR	NR	NR
PREDICT	CZP 200 mg Q2W +	368 (194 TNFi-	To week 12 303 (82.1)	Prior to week 12 66 (17.9)	Total full analysis set	NA	NA	NA	NA

Trial	Arms	Randomised, n	Completed double-blind phase, n (%)	Discontinued prior to end of double-blind phase, n (%)	Analysis sets, n (%)	Entered OLE phase	Completed OLE, n (%)	Discontinued prior to end of OLE, n	Analysis sets, n (%)
	DMARDs: RAPID3	experienced)	To week 52 187 (50.7)	AE = 13 (3.5) Lack of efficacy = 22 (6.0) Protocol violation = 9 (2.4) LtFU = 6 (1.6) WoC = 6 (1.6) Other unspecified = 10 (2.7) Prior to week 52 182 (49.3) AE = 29 (7.9) Lack of efficacy = 86 (23.3) Protocol violation = 9 (2.4) LtFU = 10 (2.7) WoC = 11 (3.0) Other unspecified = 37 (10.0)	for both groups n=733 (407 TNFi-experienced) Data reported for CZP treatment subgroups			
	CZP 200 mg Q2W + DMARDs: CDAI	365 (213 TNFi-experienced)	To week 12 314 (85.6) To week 52 192 (52.3)	Prior to week 12 53 (14.4) AE = 19 (5.2) Lack of efficacy = 20 (5.4) Protocol violation = 4 (1.1) LtFU = 3 (0.8) WoC = 4 (1.1) Other unspecified = 3 (0.8) Prior to week 52 175 (47.7) AE = 43 (11.7) Lack of efficacy = 74 (20.2) Protocol violation = 5 (1.4) LtFU = 8 (2.2) WoC = 27 (7.4) Other unspecified = 18 (4.9)	Data reported for CZP treatment subgroups	NA	NA	NA	NA
SWITCH	CZP 200 mg Q2W + DMARDs	27 (all TNFi-experienced)	To week 12 27 (100)	Prior to week 12 0	27	27	To week 24 27 (100)	Prior to week 24 0	22
	PBO Q2W + DMARDs	10 (all TNFi-experienced)	To week 12 10 (100)	Prior to week 12 0	10	→ CZP 200 mg Q2W + DMARDs 10 (100)	To week 24 9 (90)	Prior to week 24 1 (10) (reason for discontinuation	8

Trial	Arms	Randomised, n	Completed double-blind phase, n (%)	Discontinued prior to end of double-blind phase, n (%)	Analysis sets, n (%)	Entered OLE phase	Completed OLE, n (%)	Discontinued prior to end of OLE, n	Analysis sets, n (%)
								unspecified)	
J-RAPID	CZP 200 mg Q2W + MTX	82 (11 TNFi-experienced)	To week 24 66 (80.5) (NR TNFi-experienced)	Prior to week 24 Scheduled withdrawal at week 16 due to lack of efficacy (no ACR20 response at week 12 and 14): 11 Withdrawn for other reasons: 5 WoC = 1 AE = 3 Lack of efficacy ¥ = 1	11	NR	NR	NR	NR
	PBO Q2W + MTX	77 (15 TNFi-experienced)	To week 24 25 (32.5) (NR TNFi-experienced)	Prior to week 24 Scheduled withdrawal at week 16 due to lack of efficacy (no ACR20 response at week 12 and 14): 45 Withdrawn for other reasons: 7 WoC = 3 AE = 2 Lack of efficacy ¥ = 2	15	NR	NR	NR	NR
HIKARI	CZP 200 mg Q2W	116 (Combination with non-MTX DMARDs∞ n=62, monotherapy n =54)	To week 24 82 (70.7)	Prior to week 24 Scheduled withdrawal at week 16 due to lack of efficacy (no ACR20 response at week 12 and 14): 24 Withdrawn for other reasons: 10 Patient's request = 1 AE = 8 Did not receive medication more than twice = 1	6	NR	NR	NR	NR
	PBO	114 (Combination with non-MTX	To week 24 18 (15.8)	Prior to week 24 Scheduled withdrawal at week 16 due to lack of efficacy (no ACR20 response	10	NR	NR	NR	NR

Trial	Arms	Randomised, n	Completed double-blind phase, n (%)	Discontinued prior to end of double-blind phase, n (%)	Analysis sets, n (%)	Entered OLE phase	Completed OLE, n (%)	Discontinued prior to end of OLE, n	Analysis sets, n (%)
		DMARDs[∞] n=65, monotherapy n =49)		at week 12 and 14): 88 Withdrawn for other reasons: 8 Patient's request = 2 AE = 2 Protocol non-compliance = 1 Lack of efficacy¥ = 2 Did not receive medication more than twice = 1					

AE = Adverse event, LtFU = Lost to follow-up, NA= not applicable, NR = not reported, OLE = open-label extension, WoC = Withdrawal of consent

[*] 1 CZP completer discontinued OLE after week 12 due to AE, did not receive study treatment in OLE and not included in OLE analysis set

** In the DOSEFLEX study, 333 patients entered CZP 200 mg Q2W + MTX run-in phase (TNFi-experienced n=178 [53.3%]). Not randomised at week 18 n= 124 (37.2%): AE = 17, Lack of efficacy = 93, Loss of efficacy n= 1, LtFU = 4, WoC = 5, Other unspecified n=4

ǂ one patient not treated in double blind phase

¥ Efficacy of treatment insufficient at times other than weeks 12 and 14. ACR20 non-responders at week 12 and week 14 withdrawn from study at week 16 and eligible to enter OLE as were study completers.

[∞] CZP in combination with non-MTX cDMARDs not approved in European Union

4.2.4 Outcomes in included CZP RCTs

Outcomes in included studies were clearly tabulated in the CS and are collated by the ERG in Table 9. ClinicalTrials.gov records for the included CZP RCTs were checked by the ERG for the presence of any listed outcomes additional to those described in Table 9; none were identified. Whilst HAQ-DI was described as a key measure of physical function, morning stiffness data were not presented in the CS for J-RAPID and HIKARI. Radiological progression / joint damage data, whilst available in the trial publications for J-RAPID[32] and HIKARI,[33] were also not discussed in the CS. The CS included some data on additional outcomes not listed in the final NICE scope[27] (e.g. impact of CZP on sleep, work productivity, household productivity): these outcomes are not summarised in the ERG report (as they were outside of the NICE final scope).[27]

Table 9: Outcomes included in CZP RCTs (adapted from CS Tables 10, 11 and 12)

Trial name	Primary outcome(s)	Secondary outcome(s)
REALISTIC	ACR20 response at 12 weeks	• ACR20 response at week 12 based on pre-specified baseline stratification factors • ACR20/50/70 response at weeks 12 and 28 • Improvements in ACR component scores and EULAR response at weeks 12 and 28 • DAS28 reduction based on DAS28 (CRP), SDAI and CDAI at weeks 12 and 28 • DAS28 remission rates at weeks 12 and 28 • TJC, SJC, HAQ-DI, CRP improvement at weeks 12 and 28 • Improvement in pain, fatigue, patient/physician global assessment of disease activity and sleep problems at week 12 • Time to sustained ACR20 response and EULAR response • Safety
DOSEFLEX	ACR20 response at 34 weeks in patients randomised at week 18	• ACR20/50/70 responses at weeks 16 and 34 • ACR component scores at weeks 16 and 34 • EULAR response rates at weeks 16 and 34 • DAS28(ESR) response, remission rates and HAQ-DI up to week 34 • CDAI, SDAI and DAS28(ESR) remission at weeks 16 and 34 • Change from baseline in CRP at week 16 and 34 • PROs (SF-36 component summaries and domain scores, fatigue, patient global assessment of disease activity and pain) at week 34 • Change in TJC and SJC at week 34 • Physician's global assessment of disease activity at week 34 • Median time to loss of ACR20 after week 18 in patients randomised at week 18 • Safety
PREDICT	CDAI and RAPID-3 scores at 12 and 52 weeks Responders (CDAI/RAPID3) at week 12 achieving DAS28(ESR) ≤ 3.2 at 52 weeks	• Modified ACR (mACR)20/50/70 response rates • mACR component scores and EULAR response rates at weeks 12, 24 and 52 • DAS28(ESR) response and remission rates by disease activity up to week 52 • CDAI, RAPID3 and DAS28(ESR) remission up to week 52 • MD-HAQ, and work and household productivity improvement up to week 52 • Safety

SWITCH	ACR20 response at 12 weeks CDAI response at 12 weeks	• CDAI response, ACR50/70 response, Low Disease Activity (DAS28 (CRP) of ≤3.2 or CDAI <10), EULAR response at weeks 12 and 24 • Change from baseline in HAQ-DI • Safety
J-RAPID	ACR20 response at 12 weeks	• ACR20 response at 24 weeks • ACR50/70 response rates • ACR core component scores incl. HAQ-DI, pain, patient and physician global assessment of disease activity, CRP and ESR • DAS28(ESR) • EULAR response • Prevention of progression of joint damage (change in van der Heijde modified Total Sharp Score) at week 24 • Morning stiffness duration • SF-36 at weeks 12 and 24
HIKARI	ACR20 response at 12 weeks	• ACR20 response at 24 weeks • ACR50/70 response rates • ACR core component scores incl. HAQ-DI, pain, CRP and ESR • mTSS at week 24 • Morning stiffness duration • DAS28(ESR) • EULAR response • SF-36 at weeks 12 and 24

4.2.5 Summary of clinical effectiveness results for CZP

The statistical significance of comparisons made between CZP and comparator treatment arms in the TNFi-experienced trial populations were frequently not reported in the CS. The ERG requested that the company provide *p*-values where these were not reported in the CS. In response (see clarification response, Question A36[28]), the company performed *post hoc* analyses for key outcomes and provided a series of *p*-values. The company noted that analyses were exploratory (i.e. nominal *p*-values) so should be treated with caution (particularly in cases where *p*-values were based on sample sizes ≤ 15% of study population) and that no conclusions can be made on the statistical significance of comparisons. The *p*-values provided by the company in response to the ERG's clarification request have been added to the ERG report data and are marked with a symbol (†).

4.2.5.1 Disease activity

Disease activity was reported in the CS in terms of ACR and EULAR responses, DAS28 and CDAI. The summary of ACR and EULAR response data, as key disease activity outcomes, has been prioritised in this ERG report. DAS28 (REALISTIC [CS Section 4.7.1.2.2], DOSEFLEX [CS Section 4.7.2.2.2], PREDICT [CS Section 4.7.3.2.2], SWITCH [CS Section 4.7.4.1.2], J-RAPID [CS Section 4.7.5.1.2], HIKARI [CS Section 4.7.5.2.2]) and CDAI (REALISTIC [CS Section 4.7.1.2.2], DOSEFLEX [CS Section 4.7.2.2.2], PREDICT [CS Section 4.7.3.2.2]) data were included in the CS but are not summarised in the ERG report. The ERG noted that the numbers included in the REALISTIC TNFi-experienced CZP monotherapy and combination treatment subgroups in the analysis of ACR and EULAR responses at week 12 did not appear to tally with the total number of TNFi-experienced patients for reasons that are unclear to the ERG (i.e. [redacted]

4.2.5.2 ACR response

ACR response data were available from all included CZP RCTs (PREDICT reported modified ACR [mACR]) and were collated by the ERG in Table 10. The modified ACR was described in Appendix 8.6.2 of the CS and was reported to differ from the standard ACR in two aspects. Firstly, tender and swollen joints were assessed in 28 joints (used in the DAS28 assessment)[redacted] Secondly, patients' assessment of physical function, global health and pain were assessed[redacted] ACR data were provided in the CS for overall trial populations and subgroups of TNFi-experienced patients. Results for TNFi-experienced subjects only are summarised in this section.

Table 10: ACR response rates in included CZP RCTs*

Trial name	Treatment Group	Treatment arms for which data extraction performed (n)	Assessment time point	% achieving ACR20 response	% achieving ACR50 response	% achieving ACR70 response
REALISTIC	Overall population	PBO (n=212)	Week 2	[redacted]	[redacted]	[redacted]
		CZP 200 mg Q2W (n=851)	Week 2	[redacted]	[redacted]	[redacted]
		PBO (n=212)	Week 6	[redacted]	[redacted]	[redacted]
		CZP 200 mg Q2W (n=851)	Week 6	[redacted]	[redacted]	[redacted]
		PBO (n=212)	Week 12	[redacted] (25.9%)	(9.9%)	(2.8%)
		CZP 200 mg Q2W (n=851)	Week 12	[redacted] (51.1%) p <0.001	(26.6%) p <0.001	(12.9%) p <0.001
		CZP 200 mg OLE (n=770)	Week 12	[redacted]	[redacted]	[redacted]
		CZP 200 mg OLE (n=770)	Week 28	[redacted]	[redacted]	[redacted]
	Overall population, (NRI), CZP monotherapy	PBO (n=69)	Week 12	[redacted]	[redacted]	[redacted]
		CZP 200 mg Q2W (n=262)	Week 12	[redacted]	[redacted]	[redacted]
		CZP 200 mg OLE (n=237)	Week 12	[redacted]	[redacted]	[redacted]
		CZP 200 mg OLE (n=237)	Week 28	[redacted]	[redacted]	[redacted]
	Overall population, (NRI), CZP+MTX	PBO (n=143)	Week 12	[redacted]	[redacted]	[redacted]
		CZP 200 mg Q2W (n=589)	Week 12	[redacted]	[redacted]	[redacted]
		CZP 200 mg OLE (n=533)	Week 12	[redacted]	[redacted]	[redacted]
		CZP 200 mg OLE (n=533)	Week 28	[redacted]	[redacted]	[redacted]
REALISTIC	TNFi-experienced	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	TNFi-experienced (NRI), CZP monotherapy	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	TNFi-experienced (NRI), CZP+MTX	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
DOSEFLEX	Overall	PBO+MTX (n=69)	Week 4	[redacted]	[redacted]	[redacted]

Trial name	Treatment Group	Treatment arms for which data extraction performed (n)	Assessment time point	% achieving ACR20 response	% achieving ACR50 response	% achieving ACR70 response
	population	CZP 200 mg Q2W+MTX (n=70)	Week 4	[redacted]	[redacted]	[redacted]
		CZP 400 mg +MTX Q4W (n=69)	Week 4	[redacted]	[redacted]	[redacted]
		PBO+MTX (n=69)	Week 12	[redacted]	[redacted]	[redacted]
		CZP 200 mg Q2W+MTX (n=70)	Week 12	[redacted]	[redacted]	[redacted]
		CZP 400 mg +MTX Q4W (n=69)	Week 12	[redacted]	[redacted]	[redacted]
		PBO+MTX (n=69)	Week 16	[redacted]	[redacted]	[redacted]
		CZP 200 mg +MTX Q2W (n=70)	Week 16	[redacted]	[redacted]	[redacted]
		CZP 400 mg +MTX Q4W (n=69)	Week 16	[redacted]	[redacted]	[redacted]
		PBO+MTX (n=69)	Week 24	[redacted]	[redacted]	[redacted]
		CZP 200 mg +MTX Q2W (n=70)	Week 24	[redacted]	[redacted]	[redacted]
		CZP 400 mg +MTX Q4W (n=69)	Week 24	[redacted]	[redacted]	[redacted]
		PBO+MTX (n=69)	Week 34	[redacted] (44.9%)	[redacted] 30.4%)	[redacted] (15.9%)
		CZP 200 mg +MTX Q2W (n=70)	Week 34	[redacted] (67.1%)	[redacted] (50.0%)	[redacted] (30.0%)
		CZP 400 mg +MTX Q4W (n=69)	Week 34	[redacted] (65.2%)	[redacted] (52.2%)	[redacted] (37.7%)
	TNFi-experienced (NRI)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	TNFi-experienced (LOCF)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
SWITCH[¥]	TNFi-experienced	PBO Q2W + cDMARDs (n=10)	Week 12	0 (0%)	0 (0%)	0 (0%)
		CZP 200 mg Q2W+cDMARDS (n=27)	Week 12	17 (61.5%) $p<0.005$	5 (19.0%)	1 (3.5%)
		PBO Q2W + cDMARDs (n=8)	Week 24	5 (62.5%) (p value NR)	3 (37.5%) (p value NR)	NR

Trial name	Treatment Group	Treatment arms for which data extraction performed (n)	Assessment time point	% achieving ACR20 response	% achieving ACR50 response	% achieving ACR70 response
		CZP 200 mg Q2W+cDMARDS (n=22)	Week 24	12 (54.5%) (*p* value NR)	6 (27.3%) (*p* value NR)	NR
HIKARI	Overall population	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	TNFi-experienced	PBO (n=10)	Week 12	[redacted]	[redacted]	[redacted]
		CZP 200 mg Q2W (n=6)	Week 12	[redacted]*	[redacted]*	[redacted]*
		PBO (n=10)	Week 24	[redacted]	[redacted]	[redacted]
		CZP 200 mg Q2W (n=6)	Week 24	[redacted]*	[redacted]*	[redacted]*
J-RAPID	TNFi-experienced	PBO +MTX (n=15)	Week 12	[redacted]	[redacted]	[redacted]
		CZP 200 mg Q2W +MTX (n=11)	Week 12	[redacted]*	[redacted]*	[redacted]*
		PBO +MTX (n=15)	Week 24	[redacted]	[redacted]	[redacted]
		CZP 200 mg Q2W +MTX (n=11)	Week 24	[redacted]*	[redacted]*	[redacted]*

* The PREDICT trial measured mACR instead of standard ACR (data are summarised in supporting ERG report text)

¥ SWITCH 24-week data were sourced from Clinicaltrials.gov. Efficacy denominators differ from N randomised (reason unclear) and are as reported in the source material

‡ The ERG noted that the numbers included in the REALISTIC TNFi-experienced CZP monotherapy and combination treatment ACR analysis subgroups did not tally with the total number of TNFi-experienced patients for reasons unclear.

† *p*-values provided in company's clarification response

All CZP RCTs provided data on ACR (mACR in the case of PREDICT). The available data are summarised below.

At week 12 of the REALISTIC trial, TNFi-experienced patients receiving CZP + MTX were [redacted] likely to achieve: ACR20 [redacted] vs PBO [redacted] ACR50 [redacted] and ACR70 [redacted] responses than those on PBO. ACR responses were [redacted] maintained to week 28 of the CZP OLE.

TNFi-experienced patients in both CZP treatment groups of DOSEFLEX [redacted] likely than those on PBO to reach ACR20, 50 and 70 responses by week 34.[35]

In their clarification response (see clarification response, Question A8[28]), the company provided pooled data from DOSEFLEX and the follow-up study DOSEFLEX II for ACR response rates to week 42. Details were not available on the method of data pooling. Data were for the TNFi-experienced population randomised to CZP from week 18. These data supported the [redacted] of ACR20/50/70 responses to week 42 in patients [redacted] receiving CZP + MTX [redacted]

In the overall study population of PREDICT, [redacted] proportions of patients with mACR50 and mACR70 responses were observed between weeks 2 and 52 in patients receiving CZP 200 mg Q2W [redacted]

Table 11: PREDICT study: mACR20/50/70 response rates in overall study population during 52 week double-blind phase (NRI) (reproduced from CS)

	CZP 200 mg Q2W [redacted]
mACR20: n (% response rate)	
Week 2	[redacted]
Week 12	[redacted]
Week 24	[redacted]
Week 52	[redacted]
mACR50: n (% response rate)	
Week 2	[redacted]
Week 12	[redacted]
Week 24	[redacted]
Week 52	[redacted]
mACR70: n (% response rate)	
Week 2	[redacted]
Week 12	[redacted]
Week 24	[redacted]
Week 52	[redacted]

[redacted]

The CS also included mACR kinetic data for TNFi-experienced patients in the PREDICT trial over 52 weeks[redacted]

[redacted]The proportion of mACR20 responders[redacted]at week 12[redacted]before falling to[redacted]by week 52. The proportion of mACR50 was highest at week 24[redacted]and lowest at week 52[redacted]The proportions of patients classed as mACR70 responders[redacted]across weeks 12[redacted]to 52[redacted]

Figure 2: PREDICT study: kinetics of mACR20/50/70 response rates in TNFi experienced population during 52-week double-blind phase (NRI) (reproduced from CS Figure 35)

mACR responses among TNF-experienced patients receiving CZP monotherapy and MTX combination treatment in PREDICT were presented in Appendix 8.9.1.1 of the CS and appeared

Patients in the SWITCH trial who were TNFi-experienced and treated with CZP + cDMARDs had more favourable ACR20, 50 and 70 responses by week 12 than those in the PBO group (ACR20 $p<0.005$). The company clarified that ACR50/70 data at week 12 and response rates at week 24 were

estimated using graph-reading software. However, the ERG identified that 24-week data for SWITCH were available on ClinicalTrials.gov. The denominators used in the reported analysis on ClinicalTrials.gov differed from the randomised total for unspecified reasons. Interestingly, analyses reported on ClinicalTrials.gov indicated that the proportions of ACR20 and ACR50 responders were actually lower in CZP-treated patients (54.5% and 27.3%) compared with PBO subjects (62.5% and 37.5%) at week 24 (*p*-values not reported [NR]) (but analyses were based on relatively small numbers of subjects).

TNFi-experienced subjects receiving CZP in the J-RAPID and HIKARI trials had [redacted] favourable ACR20, 50 and 70 responses at weeks 12 and 24 than those on PBO, although these analyses were also based on relatively small numbers of patients

In their clarification response (see clarification response, Question A6[28]), the company provided limited ACR20 data for the small TNFi-experienced population in the CZP RCT RA0025 (Kang *et al.* 2012[39]). [redacted] proportion of CZP+MTX patients achieved an ACR20 response compared with PBO+MTX.

Table 12: RA0025 study (Kang *et al.* 2012[39]): ACR20 response rates at Week 24 (TNFi experienced population, non-responder imputation (NRI)

	PBO + MTX [redacted]	CZP 200 mg Q2W + MTX [redacted]
ACR20 Response, n (%)		
Week 24	[redacted]	[redacted]
[redacted]		

Pairwise meta-analyses were performed separately for each ACR20/50/70 responses at 3 months for CZP+MTX vs PBO+MTX and results were provided in Section 4.9 and Appendix 8.11 of the CS. A forest plot was presented giving the sample estimates of treatment effect from each study and the pooled effect from fixed effects and random effects models for ACR, EULAR and DAS28.

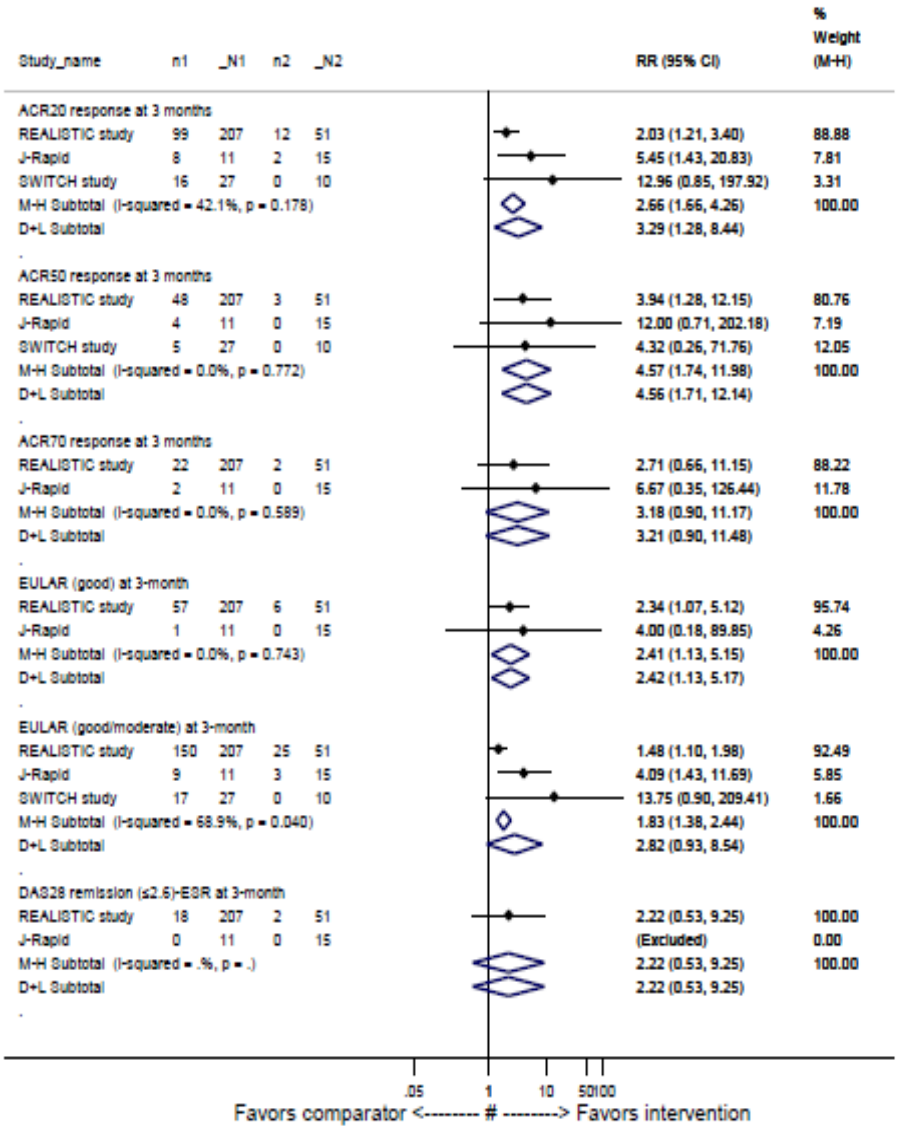
Classical fixed effect and random effects models were used to meta-analyse data from either two or three studies. It is typically recommended that a classical random effects meta-analysis should include at least five studies to estimate the between-study standard deviation; the DerSimonian and Laird estimate of the between-study standard deviation over-estimates the between-study standard deviation on average and the bias can be substantial when the number of studies is small. Such bias would also affect the estimate of I^2, thereby making it difficult to determine whether large values reflect genuine heterogeneity or whether they are simply a consequence of the bias.

The results of the fixed effect meta-analyses can be interpreted as answering the question, "Did the treatments have an effect in the studies included in the analysis?" In the CS, the results of the random effects meta-analyses may be interpreted as providing an approximately upper and lower bound to the question, "Will the treatments have an effect when given to future patients?" because of the bias associated with the DerSimonian and Laird estimate of the between-study standard deviation.

The ERG considers it inappropriate to assert that because the estimates of the treatment effects from the different analyses were all in the same direction the heterogeneity can be ignored as stated in the CS; in the presence of heterogeneity the treatment effect from a random effects model does not represent the treatment effect in any specific patient population.

The ERG would like to have seen results from a Bayesian random effects meta-analysis incorporating reasonable prior beliefs for the between-study standard deviation in the form of weakly informative prior information.

Figure 3: Direct meta-analysis results for CZP + MTX versus MTX (fixed effects model): (reproduced from CS Figure 42)



Note: D-L represents random-effects results calculated using the DerSimonian and Laird model; M-H represents fixed-effects results calculated using the Mantel-Haenszel model;
N1 = Number of patients randomised in the CZP + MTX arm; n1 = Number of patients with response at 3-month in the CZP + MTX arm; N2 = Number of patients randomised in the MTX arm; n2 = Number of patients with response at 3-month in the MTX arm; For the direction of meta-analyses results, intervention represents CZP + MTX and comparator represents MTX alone

Data for ACR response in TNFi-experienced patients receiving CZP as monotherapy were available for two CZP RCTs (REALISTIC and HIKARI).

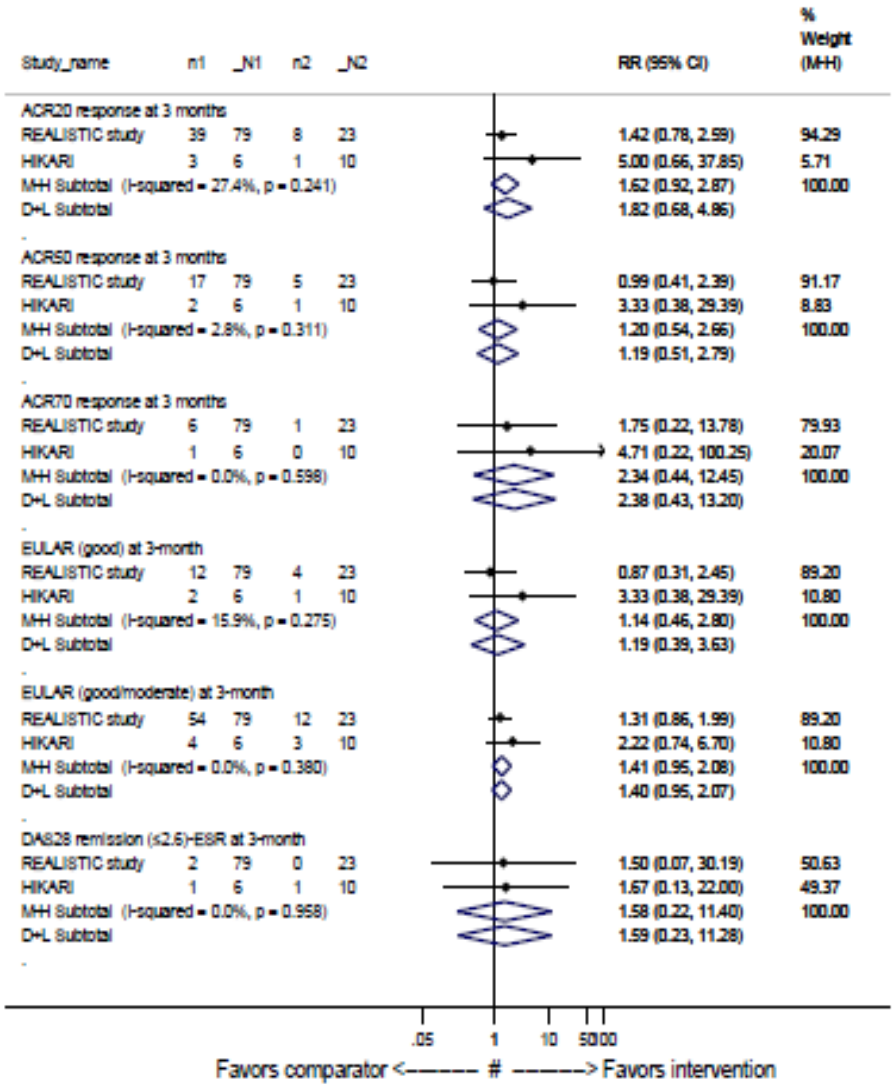
Classical meta-analysis results indicated more favourable ACR20/50/70 responses at 3 months for patients receiving CZP + MTX compared with PBO.

Trial data were also analysed for the use of CZP in monotherapy. Larger proportions of TNFi-experienced participants in the REALISTIC trial receiving CZP monotherapy reached ACR20 and ACR70 responses at week 12 compared with those on PBO Similar proportions of CZP and PBO participants reached an ACR50 response. ACR responses were maintained at similar magnitudes to week 28 of the CZP OLE

Patients in the HIKARI trial who were TNFi-experienced and being treated with CZP monotherapy were more to be responders in terms of ACR20, ACR50 and ACR70 at weeks 12 and 24 compared with the PBO group.

ACR response data from TNFi-experienced patients in REALISTIC and HIKARI were meta-analysed to compare CZP monotherapy with PBO (Figure 4). Data suggested a more favourable ACR response for CZP monotherapy patients compared with PBO, although the results were inconclusive.

Figure 4: **Direct meta-analysis results for CZP monotherapy versus PBO (fixed effects model): (reproduced from CS Figure 43)**



Note: D-L represents random-effects results calculated using the DerSimonian and Laird model; M-H represents fixed-effects results calculated using the Mantel-Haenszel model;
N1 = Number of patients randomised in the CZP monotherapy arm; n1 = Number of patients with response at 3 month in the CZP monotherapy arm; N2 = Number of patients randomised in the PBO arm; n2 = Number of patients with response at 3 month in the PBO arm; For the direction of meta-analyses results, Intervention represents CZP monotherapy and comparator represents PBO

4.2.6 EULAR response

EULAR response data were presented in the CS for all included CZP RCTs. EULAR good or moderate response week 24 data were available for the SWITCH trial in ClinicalTrials.gov. Data were summarised by the ERG and presented in Table 13.

Table 13: EULAR response rates in included CZP RCTs

Trial name / Author, year	Treatment Group	Treatment arms for which data extraction performed (n)	Assessment time point	% achieving EULAR response Good	% achieving EULAR response Moderate	% achieving EULAR response None
REALISTIC [34]	TNFi-experienced (LOCF)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	TNFi-experienced (LOCF)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	TNFi-experienced (LOCF) monotherapy	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	TNFi-experienced (LOCF) +MTX	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
DOSEFLEX [35]	TNFi-experienced (LOCF)**	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
PREDICT [36]	TNFi-experienced	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Trial name / Author, year	Treatment Group	Treatment arms for which data extraction performed (n)	Assessment time point	% achieving EULAR response Good	% achieving EULAR response Moderate	% achieving EULAR response None
	(LOCF)	[redacted]	[redacted]	[redacted]*	[redacted]*	[redacted]*
SWITCH	TNFi-experienced	PBO Q2W + DMARDs	[redacted]	NR*	NR*	NR
		CZP 200 mg Q2W + DMARDs		NR*	NR*	NR
		PBO Q2W + DMARDs		NR*	NR*	NR
		CZP 200 mg Q2W + DMARDs		NR*	NR*	NR
J-RAPID[32]	TNFi-experienced (LOCF)	[redacted]	[redacted]	[redacted]*	[redacted]*	[redacted]*
HIKARI[33]	TNFi-experienced (LOCF)	[redacted]	[redacted]	[redacted]*	[redacted]*	[redacted]*

NR = not reported

* EULAR combined good or moderate response data were available for SWITCH in Clinicaltrials.gov and are presented in Table 14

[redacted]

In their clarification response (see clarification response, Question A36[28]), the company stated [redacted] of CZP 200 mg Q2W patients had a good or moderate EULAR response versus [redacted] in PBO patients [redacted] TNFi-experienced patients receiving CZP + MTX in REALISTIC were [redacted] likely to achieve good or moderate EULAR responses [redacted] at week 12 than those on PBO [redacted] than PBO subjects [redacted] to have a good or moderate EULAR response at 12 weeks [redacted] . The proportion of patients achieving a good EULAR response [redacted] for patients on CZP + MTX than CZP monotherapy [redacted] In their clarification response (see clarification response, Question A36[28]), the company stated that in the TNFi-experienced population, CZP 200 mg Q2W monotherapy-treated patients were [redacted] to reach a good or moderate EULAR response at week 12 than PBO patients [redacted].

In DOSEFLEX [redacted] of TNFi-experienced patients in the CZP + MTX and PBO + MTX treatment arms achieved EULAR good or moderate response rates at [redacted] However, by week 34, TNFi-experienced patients treated with CZP 200 mg Q2W + MTX [redacted] and CZP 400 mg Q4W + MTX [redacted] to reach a good EULAR response than those receiving PBO + MTX [redacted] In their clarification response (see clarification response, question A36[28]), the company reported that CZP-treated patients were [redacted] to be EULAR good responders [redacted] The company provided in their clarification response pooled data from DOSEFLEX AND DOSEFLEX II for EULAR response to week 42 in the TNFi-experienced population randomised to CZP + MTX from week 18. The method of data pooling was not described. The proportions of patients achieving good or moderate responses [redacted]

The proportions of TNFi-experienced patients treated with CZP reaching moderate or good EULAR responses in the PREDICT trial were [redacted] maintained between weeks 12 and 52 [redacted]

TNFi-experienced patients receiving CZP + MTX in the J-RAPID trial were [redacted] likely to reach moderate [redacted] or good [redacted] EULAR responses at week 12 than PBO group subjects [redacted]

In the HIKARI trial [redacted] of CZP-treated TNFi-experienced patients achieved good [redacted] or moderate [redacted] EULAR responses at [redacted] CZP benefits were maintained at 24 weeks for good EULAR response [redacted] but were [redacted] between groups for moderate EULAR response [redacted]

EULAR good or moderate response data at weeks 12 and 24 were available for SWITCH in ClinicalTrials.gov. The reason for the different denominators in the analyses for weeks 12 and 24 was unclear. Whilst CZP-treated patients demonstrated a much more favourable EULAR response at week 12, results at week 24 were roughly comparable between CZP and PBO treatment groups; *p*-values were not reported.

Table 14: EULAR good or moderate response in SWITCH (source: Clinicaltrials.gov)

Treatment arm	Time of assessment	Number (%) of patients achieving a EULAR good or moderate response*
CZP 200 mg Q2W + DMARDs	Week 12	17/26 (65.4%)
PBO Q2W + DMARDs	Week 12	0/9 (0.0%)
CZP 200 mg Q2W + DMARDs	Week 24	17/22 (77.3%)
PBO Q2W + DMARDs	Week 24	6/8 (75.0%)

* *p*-values not reported

EULAR data at 3 months were meta-analysed to compare the effects of CZP (both in combination with MTX and as monotherapy) with PBO. The forest plots indicated [redacted]

4.2.7 Physical function

Four of the included CZP RCTs reported physical function in terms of Health Assessment Questionnaire Disability Index (HAQ-DI) scores (or MD-HAQ/M-HAQ in the case of PREDICT). Whilst HAQ-DI outcome data were collected at weeks 12 and 24 in the J-RAPID[32] and HIKARI[33] studies, these data were not presented in the CS.

In the TNFi-experienced patients in REALISTIC, the mean HAQ-DI scores in both CZP [redacted] and PBO [redacted]) treatment groups showed [redacted] to week 12 [redacted] The CS states that for HAQ-DI score there was a [redacted] interaction between treatment and TNFi experience [redacted]

Figure 5: REALISTIC study: HAQ-DI to Week 12 for TNFi experienced population during 12 week double-blind phase (OLE scores in inset table) (LOCF) (reproduced from CS Figure 24)

It was also reported in the CS that HAQ-DI scores within the REALISTIC RCT were At week 12 of REALISTIC, TNFi-experienced patients treated with CZP monotherapy had a For TNFi-experienced patients treated with CZP+MTX, the HAQ-DI mean change from baseline to week 12 was

All treatment groups in the DOSEFLEX trial showed HAQ-DI during the 16 week open label run-in period

However, differences between groups were evident following randomisation into the double-blind phase, with maintenance of physical function to week 34 in patients receiving CZP but worsening in PBO group patients

Figure 6: DOSEFLEX study: kinetics of HAQ-DI score for TNFi experienced population during the first 34 weeks of study (LOCF) (reproduced from CS Figure 32)

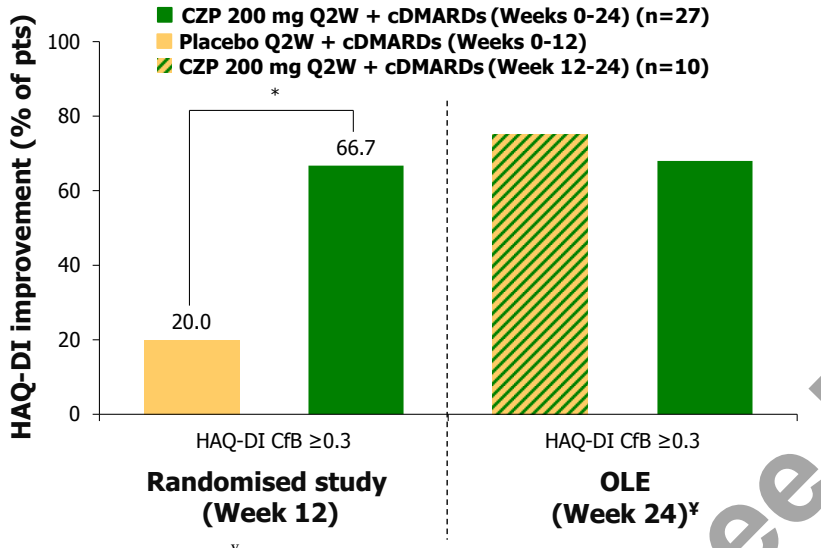
In response to a request for clarification, the company provided pooled data from DOSEFLEX and DOSEFLEX II for HAQ-DI scores to week 42. Data were for the TNFi-experienced population randomised to CZP + MTX from week 18. At week 42, patients (n=64) receiving CZP + MTX experienced a [redacted] from baseline in HAQ score.

MD-HAQ was measured in the PREDICT trial. The MD-HAQ was derived from HAQ-DI and was based on seven domains (described in Appendix 8.6.2 of the CS). [redacted] in MD-HAQ physical function were observed to week 12 and maintained to week 52 [redacted] The response was stated to be comparable between the TNFi-experienced population and CZP + MTX combination and monotherapy groups [redacted]

Figure 7: PREDICT study: kinetics of MD-HAQ global score for TNFi experienced population during 52 week double-blind phase (LOCF) (reproduced from CS Figure 38)

The proportion of patients with HAQ-DI improvement (defined as a reduction in HAQ-DI of ≥0.3) was measured in the SWITCH trial (see Figure 8).[37] A greater proportion of CZP patients (66.7%) demonstrated this improvement compared with PBO group patients (20%) (p =0.046). In the OLE, PBO group patients who moved to CZP treatment experienced HAQ-DI improvement by week 24.

Figure 8: SWITCH study: HAQ-DI improvement for TNFi experienced (overall) population at Weeks 12 and 24 of OLE phase (reproduced from CS Figure 41)



*p=0.046 CZP vs PBO. ¥Data for Week 24 have been re-drawn from the manuscript. Data point values are not available. Figure adapted from Schiff *et al.* 2014[37]

4.2.8 Joint damage/radiological progression

No data for the outcomes of joint damage or radiological progression were included in the CS. However, data on inhibition of joint structural damage were reported in the published articles for both J-RAPID and HIKARI. Both trials included radiographic assessments at baseline and week 24 or at discontinuation using modified Total Sharp Score (mTSS). Due to time constraints these data have not been extracted by the ERG.

4.2.9 Pain

Pain was listed as a secondary outcome in the REALISTIC, DOSEFLEX, J-RAPID and HIKARI RCTs. However, pain as an outcome was only covered briefly in the CS.

Bodily pain as a domain of the SF-36 was reported for the DOSEFLEX study (CS Table 20). Both PBO and CZP (200 mg Q2W and 400 mg every 4 weeks Q4W) treatment groups experienced [redacted] in bodily pain score between baseline and week 16 (end of open-label CZP run-in phase).During the double-blind period (to week 34), PBO group patients had [redacted] in SF-36 bodily pain score, whilst scores were maintained in patients receiving CZP.

Data for pain as a component of the ACR response were presented for TNFi-experienced patients in REALISTIC (Appendix 8.7.3.1 of the CS and updated in clarification response Question A36[28]) and DOSEFLEX (Appendix 8.8.3.1 of the CS).

Subjects receiving CZP 200 mg Q2W in REALISTIC demonstrated a [redacted] in patient's assessment of arthritis pain score from baseline to week 12 than placebo group patients [redacted] were maintained between weeks 12 and 28 in patients receiving CZP 200 mg Q2W in the OLE [redacted] in patient's assessment of arthritis pain score from baseline to week 12were evident in TNFi-experienced patients in REALISTIC in both CZP monotherapy [redacted] and combination with MTX subgroups [redacted]

In the open-label CZP run-in period to week 16 of the DOSEFLEX study, reductions in patient's assessment of arthritis pain as an ACR component were observed for [redacted] However, during the randomised controlled period (to week 34), PBO plus MTX group patients experienced [redacted]

Data on patient's assessment of arthritis pain were reported in the trial publications for J-RAPID (Yamamoto 2014a[32]) and HIKARI (Yamamoto 2014b[33]) but were not presented in the CS.

4.2.10 Mortality

Data relating to deaths among CZP RCT participants are presented in the safety section of this report.

4.2.11 Fatigue

The impact of CZP on fatigue was assessed in the REALISTIC and DOSEFLEX studies.

In the REALISTIC RCT patients were assessed using the Fatigue Assessment Scale. Subjects were required to answer the following question: "Please rate your fatigue (weariness, tiredness) during the past 7 days, on a scale of 0 to 10" where 0 is 'No Fatigue' and 10 is 'Fatigue as bad as you can imagine'. Clinically meaningful reductions in fatigue (defined as ≥1-point improvement) [redacted]

Figure 9: **REALISTIC study: Patient's Assessment of Fatigue in TNFi experienced population during 12 week double-blind phase (OLE scores in inset table) (LOCF) (reproduced from CS Figure 22)**

Fatigue data were broken down by TNFi-experienced monotherapy and combination with MTX subgroups (Appendix 8.7.1.7 of the CS, see clarification response Question A36[28]). A slightly greater numerical reduction in fatigue was observed at week 12 for CZP-treated patients compared with PBO in the CZP monotherapy subgroup. The difference in fatigue between treatment arms was in the CZP combination subgroup, at patients receiving CZP + MTX and for PBO group subjects

The DOSEFLEX trial utilised the 10-point Fatigue Assessment Scale. Fatigue improved to week 16 across all treatment groups during the open-label CZP run-in Whilst fatigue improvements were maintained to week 34 of the randomised controlled period in both CZP treatment groups, the PBO group experienced a worsening of fatigue.

Figure 10: DOSEFLEX study: kinetics of Patient's Assessment of Fatigue scores in TNFi experienced population during the first 34 weeks of study (LOCF) (reproduced from CS Figure 31)

4.2.12 Extra-articular manifestations of disease

Whilst extra-articular manifestations of disease was listed as an outcome of interest in the final NICE scope,[27] no data were presented in the CS.

4.2.13 Health-related quality of life

SF-36 data were only available in the CS for the DOSEFLEX trial and are presented in

Table 15. The company provided *p*-values[†] from exploratory *post hoc* analyses (see clarification response, Question A36[28]). Table 12 of the CS indicated that SF-36 was measured in J-RAPID and HIKARI. SF-36 outcome data at weeks 12 and 24 were included in the trial publications for J-RAPID[32] and HIKARI[33] but were not presented in the CS. Whilst EQ-5D data for the PREDICT trial were not included in the clinical effectiveness section of the CS, data were presented within the description of the cost-effectiveness model (CS Section 5.4.1).

Table 15**:** SF-36 domain scores for TNFi-experienced population in DOSEFLEX (LOCF) (reproduced from CS Table 20)

SF-36 Domain	**Mean score (SD)**		
Physical Function	PBO + MTX (n=29)	CZP 200 mg Q2W + MTX (n=43)	CZP 400 mg Q4W + MTX(n=39)
Week 0	[redacted]	[redacted]	[redacted]
Week 16	[redacted]	[redacted]	[redacted]
Week 34	[redacted]	[redacted]	[redacted]
Role Physical			
Week 0	[redacted]	[redacted]	[redacted]
Week 16	[redacted]	[redacted]	[redacted]
Week 34	[redacted]	[redacted]	[redacted]
Bodily Pain			
Week 0	[redacted]	[redacted]	[redacted]
Week 16	[redacted]	[redacted]	[redacted]
Week 34	[redacted]	[redacted]	[redacted]
General Health			
Week 0	[redacted]	[redacted]	[redacted]
Week 16	[redacted]	[redacted]	[redacted]
Week 34	[redacted]	[redacted]	[redacted]
Vitality, mean			
Week 0	[redacted]	[redacted]	[redacted]
Week 16	[redacted]	[redacted]	[redacted]
Week 34	[redacted]	[redacted]	[redacted]
Social Functioning			
Week 0	[redacted]	[redacted]	[redacted]
Week 16	[redacted]	[redacted]	[redacted]
Week 34	[redacted]	[redacted]	[redacted]
Role Emotional			
Week 0	[redacted]	[redacted]	[redacted]
Week 16	[redacted]	[redacted]	[redacted]
Week 34	[redacted]	[redacted]	[redacted]
Mental Health			
Week 0	[redacted]	[redacted]	[redacted]
Week 16	[redacted]	[redacted]	[redacted]
Week 34	[redacted]	[redacted]	[redacted]

[redacted]

From weeks 0 to 16 (the end of the open-label run-in phase, in which all groups received CZP), all treatment arms experienced [redacted] From weeks 16 to 34 (i.e. during the randomised PBO-controlled portion of DOSEFLEX), participants in the PBO group [redacted] some of the benefits they had gained in all SF-36 domains [redacted] For both CZP treatment groups [redacted] in SF-36 were largely maintained or showed slight improvements in score between weeks 16 and 34.

4.2.14 Subgroup analysis

The final NICE scope[27] had indicated that (evidence permitting) the appraisal should consider the following patient subgroups:

i) having primary or secondary failure of response to a first TNFi,

ii) having seronegative or seropositive antibody status

No subgroup analyses were included in the submission (Section 4.8 of the CS).

4.2.15 Non-randomised and non-controlled evidence

It was stated in the CS that the ARTIS CZP study (Chatzidionysiou *et al.* 2015[49]) was identified during searches but was excluded from the review as it was a single-arm uncontrolled study. However, the ARTIS study was included in the submission on the company's justification that it provided supporting data on the efficacy of CZP in patients who had failed ≥ 1 TNFi.

ARTIS was an observational registry study based in Sweden. Data were collected for RA patients who initiated treatment with CZP (n=945) during the study period (October 2009 to June 2013). These patients were categorised into three subgroups: i) TNFi-naïve (n=540); ii) 1 prior TNFi (n=215); iii) ≥ 2 prior TNFi (n=190) (with previous TNFi treatment discontinued for ineffectiveness, intolerance, other unspecified). Outcomes were measured at 3 and 6 months after start of CZP treatment and included DAS28 response and score, HAQ score, EULAR response, and survival.

Baseline characteristics for participants of the ARTIS study are presented in

Table 16.

Table 16: ARTIS study: characteristics of participants across different groups (reproduced from CS Table 51)

	Overall cohort (n=945)	TNFi-naïve (n=540)	1 prior TNFi (n=215)	≥2 prior TNFis (n=190)
Mean age (SD), years	56.4 (13.8)	55.7 (13.9)	57.7 (13.7)	57.1 (13.6)
Female, n (%)*	75.2%	72.2%	75.3%	83.7%
Mean disease duration, years (SD)**	[937] 9.1 (3.6–17.7)	[535] 6 (2–12.8)	[213] 10.9 (5.6–18.9)	[188] 15 (9.9–23.7)
Use of DMARDs, % Yes***	65.4%	70.2%	63.7%	53.7%
DAS28(ESR) score, mean (SD)	[753] 4.6 (1.4)	[447] 4.6 (1.4)	[159] 4.6 (1.4)	[147] 5.0 (1.5)****
HAQ score, mean (SD)	[820] 1.1 (0.7)	[474] 1.0 (0.6)	[181] 1.1 (0.6)	[165] 1.4 (0.7)*****

SD: standard deviation;
*p=0.002 (0 prior TNFi vs 2 prior TNFi); p=0.04 (1 prior TNFi vs 2 prior TNFi)
**p< 0.0001 between all groups pairwise
***p=0.08 (0 prior TNFi vs 1 prior TNFi); p<0.0001 (0 prior TNFi vs 2 prior TNFi); p=0.04 (1 prior TNFi vs 2 prior TNFi)
****p=0.01 vs TNFi-naïve and p=0.04 vs 1 prior TNFi
*****p<0.0001 vs TNFi-naïve and p=0.003 vs 1 prior TNFi
N numbers for group presented in square brackets where they differ from the column heading
Adapted from Chatzidionysiou *et al.* 2015[49]

The company noted some imbalances in baseline characteristics in the patient subgroups and that the ARTIS authors attempted to address the problem by adjusting for significantly differing variables using Cox regression analysis. Missing data were also highlighted by the company as an issue with ARTIS, with for example, only 70% of subjects with DAS28 data at baseline having DAS28 data at 6 months. Missing data were handled using non-responder imputation.

A total of 953 participants who began treatment with CZP were analysed, of which 753 patients had DAS28 scores and baseline, and 513 had DAS28 scores at 6 months' follow-up.

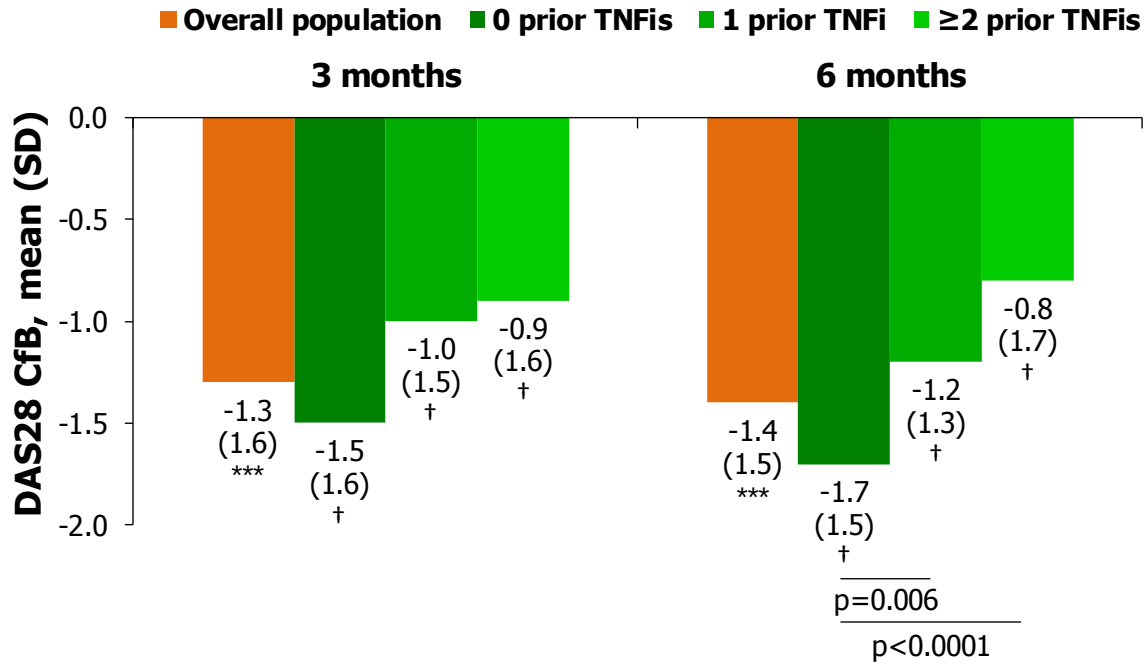
The company presented the full quality assessment for ARTIS in CS Appendix 8.10. Issues highlighted by the company included the inherent biases of the study design, resulting from lack of blinding and comparators, confounding and selection bias.

Disease activity (DAS28 ESR scores) and HAQ scores were reported. The company noted significant benefits of CZP treatment in both TNFi-naïve and TNFi-experienced patients in terms of DAS28 (p<0.0001,

Figure 11) and HAQ (p<0.0001,

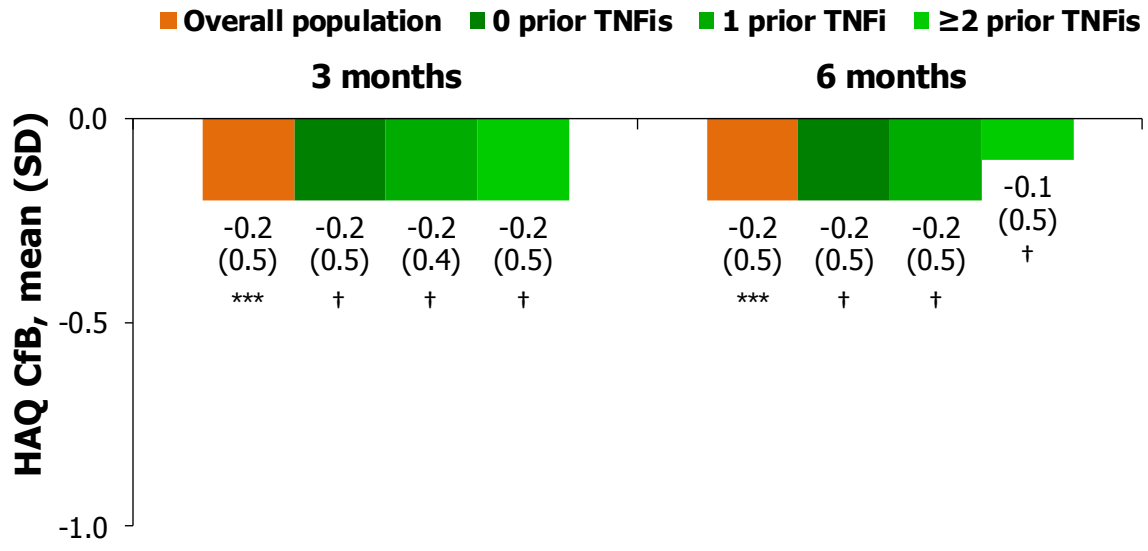
Figure 12) at 3 and 6 months following start of therapy. Changes in DAS28 from baseline were larger for TNFi-naïve patients and HAQ response were consistent across groups.

Figure 11: ARTIS study: mean change from baseline in DAS28 scores after 3 and 6 months treatment with CZP (NRI)[‡] (reproduced from CS Figure 55)



***p<0.0001
Components utilised for score are listed as TJC, SJC, general health and ESR by Chatzidionysiou *et al*.[49]
†significant reduction; p-value not reported
‡166 patients with missing follow-up data were imputed as non-responders at 6 months based on available information on treatment discontinuation.
Group sizes at 3 months were n=321, n=197, n=67 and n=57, respectively
Group sizes at 6 months were n=440, n=267, n=89 and n=84, respectively
Adapted from Chatzidionysiou *et al*. 2015[49]

Figure 12: ARTIS study: mean change from baseline in HAQ scores after 3 and 6 months treatment with CZP (Censoring)[‡] (reproduced from CS Figure 56)



***p<0.0001
†significant reduction; p-value not reported
‡Patients who had no follow-up visit were censored.
Group sizes at 3 months were n=382, n=229, n=86 and n=67, respectively
Group sizes at 6 months were n=472, n=273, n=100 and n=93, respectively
Adapted from Chatzidionysiou *et al*. 2015[49]

4.2.16 Review of safety evidence for CZP

Safety data for CZP were presented in the CS (see Section 4.12) and were largely based on the overall populations from four (REALISTIC, DOSEFLEX, PREDICT and SWITCH) of the six included CZP RCTs. The ERG requested clarification from the company as to why the CZP trials J-RAPID and HIKARI were not included in the safety review. The company stated (see clarification response, Question A16[28]) that they considered there to be some differences in how adverse events data were collected in Japanese trial populations which would limit the comparison of safety data in Japanese versus non-Japanese populations. However, safety data for J-RAPID and HIKARI were provided by the company.

A pooled analysis[50] (assessing CZP safety from data from relevant RCTs and OLEs) was also included in the CS alongside three additional studies describing the safety of CZP (Yun *et al.* 2016[51], Simard *et al.* 2011,[52] Curtis *et al.* 2015[36]).

The ERG asked the company to state whether the occurrence of any AEs in the included CZP RCTs were statistically significant for patients receiving CZP compared with PBO in (i) the overall population, and (ii) the anti-TNF experienced population. The company responded (see clarification response, Question A15[28]) that (as the included RCTs were not powered for the detection of differences in safety), statistical comparisons were not provided for safety data for the overall population or TNFi-experienced population.

AE data from the overall trial populations for REALISTIC, DOSEFLEX and SWITCH were collated by the ERG and tabulated (Table 17). As PREDICT does not have a PBO group, this trial was not included in the ERG's table.

Safety data are summarised in the following text (see Table 17 to

Table 23) with emphasis placed on key safety considerations highlighted in the CZP SmPC.[29]

4.2.16.1 Any adverse events

Slightly more patients receiving CZP 200 mg Q2W (67.5%) in the REALISTIC trial experienced an adverse event compared with those on PBO (61.7%). Among TNFi-experienced patients in REALISTIC, more patients receiving CZP 200 mg Q2W experienced an AE (68.1%) than PBO (50.0%). The occurrence of any AE among participants in the randomised controlled phase of DOSEFLEX was broadly consistent across both CZP treatment (62.9% of CZP 200 mg Q2W and 60.9% of CZP 400 mg Q4W) and PBO (62.3%) groups. A greater proportion of the CZP 200 mg Q2W group (59.3%) in SWITCH had an AE versus 40.0% in the PBO treatment arm.

4.2.16.2 Serious adverse events

The proportions of patients experiencing SAEs in the REALISTIC trial were similar (CZP 200 mg Q2W 6.1%; PBO 5.7%). Specific reported reasons for SAEs in the REALISTIC CZP groups included serious infections (CZP 200 mg Q2W 4.3%), cardiac disorders (CZP 200 mg Q2W, 1.4%), and musculoskeletal/connective tissue disorders (CZP 200 mg Q2W 2.9%, CZP 400 mg Q4W, 1.4%). For TNFi-experienced REALISTIC patients, slightly more patients on CZP 200 mg Q2W (7.9%) reported SAEs than those on PBO (5.0%) (reasons for SAEs unspecified). Both CZP treatment groups in the DOSEFLEX randomised controlled period had greater proportions of SAEs (CZP 200 mg Q2W 7.1%, CZP 400 mg Q4W 2.9%) than those in the PBO group (0.0%). No serious adverse events (SAEs) were reported for SWITCH participants.

4.2.16.3 Adverse events leading to withdrawal/discontinuation of treatment

In REALISTIC, 10% of PBO group subjects had AEs leading to withdrawal, with no cases in the CZP 200 mg Q2W arm. Of the TNFi-experienced subjects in REALISTIC, slightly more CZP 200 mg Q2W-treated patients (5.4%) had AEs leading to withdrawal than PBO group participants (2.5%). However, more patients in the REALISTIC CZP treatment arms had AEs that led to permanent discontinuation (CZP 200 mg Q2W 5.7%, CZP 400 mg Q4W 1.4%) compared with none in the PBO group. In the randomised controlled phase of DOSEFLEX, patients in all groups experienced AEs leading to withdrawal (CZP 200 mg Q2W 17.1%; CZP 400 mg Q4W 8.7%; PBO 11.6%).

4.2.16.4 Infections

Infections were more common for patients receiving CZP 200 mg Q2W (29.0%) in REALISTIC compared with PBO (23.0%). A greater proportion of TNFi-experienced patients in REALISTIC who were treated with CZP 200 mg Q2W had infections/infestations (29.3%) versus the PBO group (23.8%). The proportions experiencing infections in the randomised controlled period of DOSEFLEX

were similar across CZP (CZP 200 mg Q2W 28.6%, CZP 400 mg Q4W 36.2%) and PBO (34.8%) treatment arms. Infections were not reported in the SWITCH trial.

4.2.16.5 Injection site reactions

Injection site reactions appeared more common for patients receiving CZP compared with PBO to the end of the 12 week double-blind phase of REALISTIC (CS Table 52). Among the overall population in REALISTIC, a greater proportion of CZP patients (CZP 200 mg Q2W→CZP 200 mg Q2W, 4.9%) had injection site reactions compared with those who switched from PBO to CZP 200 mg Q2W (1.8%).

In their clarification responses (Question A16[28]), the company provided further safety data for J-RAPID. Administration site reactions (2.5% and 0.0%), injection site erythema (0.8% and 0.0%), injection site hematoma (0.4% and 0.0%), injection site haemorrhage (0.0% and 1.3%), injection site mass (0.4% and 0.0%) and injection site reaction (0.4% and 1.3%) were reported across all CZP-treated groups and the PBO group, respectively. The company stated that the majority of reactions were mild.

4.2.16.6 Malignancy

Cases of malignancy were not reported in the CS for REALISTIC, DOSEFLEX or SWITCH, although new cases may not be anticipated to have occurred given the short durations of the included trials. No cases of malignancy were reported to have occurred to week 24 in CZP or PBO treatment arms of J-RAPID (see clarification response, Question A16[28]). In HIKARI, one case of malignant disease was reported in the PBO group (see clarification response, Question A16[28]).

4.2.16.7 Additional safety studies included in CS

Data were included in the CS from the study by Yun *et al.*[51] comparing biologics in terms of first hospitalised infections (

Table 23), type of infection, number of hospitalised infections and mortality associated with biologics (Table 24). Supporting data were also summarised in the CS from the studies by Simard *et al.* 2011[52] and Curtis *et al.* 2015.[36]

Table 17: Adverse events in overall study populations for CZP RCTs[¥] (collated by ERG)

AEs	DOSEFLEX			REALISTIC		SWITCH	
	CZP 200 mg Q2W (n=70)	**CZP 400 mg Q4W (n=70)**	**PBO (n=69)**	**CZP 200 mg Q2W (n=846)**	**PBO (n=209)**	**CZP 200 mg Q2W + cDMARDs → CZP 200 mg Q2W + cDMARDs (n=27)**	**PBO + cDMARDs → CZP 200 mg Q2W + cDMARDs (n=10)***
AEs, incidence rate/100 PY (n, patient %)							
Any AEs	312.1 (44, 62.9)	299.9 (42, 60.9)	323.6 (43, 62.3)	522.1 (571, 67.5)	483.2 (129, 61.7)	16 (59.3)	4 (40.0)
Mild	-	-	-	248 (29.3)	56 (26.8)	7 (25.9)	3 (30.0)
Moderate	-	-	-	257 (30.4)	58 (27.8)	9 (33.3)	1 (10)
Severe	-	-	-	66 (7.8)	15 (7.2)	0 (0)	0 (0)
Infection and infestations	104.9 (20, 28.6)	132.4 (25, 36.2)	136.2 (24, 34.8)	143.9 (245, 29.0)	112.5 (48, 23.0)	-	-
Lower respiratory tract infections	-	-	-	3.5 (7, 0.8)	2.1 (1, 0.5)	-	-
Upper respiratory tract infections	23 (5, 7.1)	36.2 (8, 11.6)	46.5 (10, 14.5)	59.3 (112, 13.2)	41.5 (19, 9.1)	5 (17.6)	0 (0)
Nasopharyngitis	4.4 (1, 1.4)	4.4 (1, 1.4)	18.4 (4, 5.8)	-	-	-	-
Sinusitis	9 (2, 2.9)	13.1 (3, 4.3)	0 (0)	-	-	-	-
Urinary tract infection	23.1 (5, 7.1)	27.6 (6, 8.7)	33.4 (7, 10.1)	2.5 (5, 0.6)	4.2 (2, 1.0)	-	-
Ear infections	0 (0)	13.3 (3, 4.3)	0 (0)	-	-	-	-
Streptococcal infections	-	-	-	0 (0)	2.1 (1, 0.5)	-	-
Musculoskeletal/connective tissue disorders	37.6 (8, 11.4)	51.4 (11, 15.9)	64.2 (13, 18.8)	-	-	-	-
Arthralgia	4.5 (1, 1.4)	22.5 (5, 7.2)	8.9 (2, 2.9)	-	-	-	-
Back pain	13.5 (3, 4.3)	0 (0)	4.4 (1, 1.4)	-	-	-	-
RA aggravation	4.4 (1, 1.4)	8.9 (2, 2.9)	27.7 (6, 8.7)	-	-	-	-
Pain in extremity	8.9 (2, 2.9)	0 (0)	13.5 (3, 4.3)	-	-	-	-
Nervous system disorders	22.8 (5, 7.1)	17.8 (4, 5.8)	4.4 (1, 1.4)	-	-	-	-
Dizziness	13.5 (3, 4.3)	0 (0)	4.4 (1, 1.4)	-	-	-	-
Headache	9 (2, 2.9)	0 (0)	0(0)	24.2 (47, 5.6)	23.5 (11, 5.3)	2 (7.4)	0 (0)

AEs	DOSEFLEX			REALISTIC		SWITCH	
	CZP 200 mg Q2W (n=70)	**CZP 400 mg Q4W (n=70)**	**PBO (n=69)**	**CZP 200 mg Q2W (n=846)**	**PBO (n=209)**	**CZP 200 mg Q2W + cDMARDs → CZP 200 mg Q2W + cDMARDs (n=27)**	**PBO + cDMARDs → CZP 200 mg Q2W + cDMARDs (n=10)***
AEs, incidence rate/100 PY (n, patient %)							
Skin/subcutaneous tissue disorders	22.7 (5, 7.1)	22.7 (5, 7.2)	22.4 (5, 7.2)	-	-	-	-
Rash	8.9 (2, 2.9)	0 (0)	4.4 (1, 1.4)	-	-	-	-
Respiratory/thoracic/mediastinal disorders	28 (6, 8.6)	4.4 (1, 1.4)	46.9 (10, 14.5)	-	-	-	-
Cough	0 (0)	0 (0)	13.4 (3, 4.3)	-	-	2 (7.4)	3 (10)
Gastrointestinal disorders	43.9 (9, 12.9)	37.9 (8, 11.6)	41.6 (9, 13)	-	-	-	-
Nausea/vomiting	13.8 (3, 4.3)	0 (0)	4.4 (1, 1.4)	21.5 (42, 5.0)	28.2 (13, 6.2	-	-
General disorders/administration site conditions	27.8 (6, 8.6)	13.3 (3, 4.3)	22.8 (5, 7.2)	-	-	-	-
Pyrexia	18.1 (4, 5.7)	0 (0)	4.4 (1, 1.4)	-	-	-	-
SAEs	23.1 (5, 7.1)	8.8 (2, 2.9)	0 (0%)	26.7 (52, 6.1)	25.8 (12, 5.7)	0 (0)	0 (0)
Serious infections	13.6 (3, 4.3)	0 (0)	0 (0)	11.1 (22, 2.6)	8.3 (4, 1.9)		
Cardiac disorders	4.5 (1, 1.4)	0 (0)	0 (0)	-	-	-	-
Musculoskeletal/connective tissue disorders	9 (2, 2.9)	4.4 (1, 1.4)	0 (0)	-	-	-	-
Respiratory/thoracic/mediastinal disorders	0 (0)	0 (0)	0 (0)	-	-	-	-
Death	0 (0)	0 (0)	0 (0)	1.0 (2, 0.2)	0 (0)	0 (0)	0 (0)
AEs leading to withdrawal	58.4 (12, 17.1)	27.5 (6, 8.7)	37.3 (8, 11.6)	-	-	0 (0)	1 (10.0)
AEs leading to permanent discontinuation	18.4 (4, 5.7)	4.4 (1, 1.4)	0 (0)	-	-	-	-
Related AEs	-	-	-	-	-	3 (11.1)	0 (0)
Rheumatoid arthropathies	-	-	-	18.8 (37, 4.4)	37.0 (17, 8.1)		

AEs	DOSEFLEX			REALISTIC		SWITCH	
	CZP 200 mg Q2W (n=70)	**CZP 400 mg Q4W (n=70)**	**PBO (n=69)**	**CZP 200 mg Q2W (n=846)**	**PBO (n=209)**	**CZP 200 mg Q2W + cDMARDs → CZP 200 mg Q2W + cDMARDs (n=27)**	**PBO + cDMARDs → CZP 200 mg Q2W + cDMARDs (n=10)***
AEs, incidence rate/100 PY (n, patient %)							
Any TEAEs	-	-	-	-	-	16 (59.3)	4 (40.0)
Serious TEAEs	-	-	-	-	-	0 (0)	0 (0)
Mild TEAEs	-	-	-	-	-	7 (25.9)	3 (30.0)
Moderate TEAEs	-	-	-	-	-	9 (33.3)	1 (10)
Severe TEAEs	-	-	-	-	-	0 (0)	0 (0)

Notes: Data represent DOSEFLEX randomised phase weeks 17 to 34, REALISTIC 12 weeks' double blind phase, and SWITCH study 24 weeks' double blind and open label phases. Data from the 54 weeks of trial for the PREDICT study did not have a PBO arm and therefore not included in table. SAEs – serious adverse events. TEAE – treatment emergent adverse event.

[¥] Safety data were not presented in the CS for J-RAPID and HIKARI

[*] PBO patients commenced treatment with CZP during the 12-week open label phase.

Table 18: REALISTIC safety data to week 12 (TNFi-experienced patients) (reproduced from CS Table 54)

Exposure and AEs	**CZP 200 mg Q2W (n=317)**	**PBO (n=80)**
AEs, n (%)		
Any AEs	216 (68.1)	40 (50.0)
Infections and infestations	93 (29.3)	19 (23.8)
SAEs	25 (7.9)	4 (5.0)
AEs leading to withdrawal	17 (5.4)	2 (2.5)

Table 19: REALISTIC safety data for OLE phase (Weeks 12 to 28, overall safety population) (reproduced from CS Table 53)

Exposure and AEs	**CZP 200 mg Q2W→ CZP 200 mg Q2W (n=770)**	**Week 12 PBO→CZP 200 mg Q2W (n=184)**
Any AEs, incidence/100 PY (n, %)	239.1 (521, 67.7)	328.9 (142, 77.2)
SAEs, incidence/100 PY (n, %)	13.0 (56, 7.3)	20.6 (21, 11.4)
Serious infections, incidence/100 PY (n, %)	4.1 (18, 2.3)	5.7 (6, 3.3)
Death, n (%)	1 (0.5)	1 (0.1)
AEs leading to withdrawal, n (%	3 (1.6)	26 (3.4)
Injection and infusion site reactions, incidence/100 PY (n, %)	8.8 (9, 4.9)	3.2 (14, 1.8)

Table 20: Safety during the 52 weeks of the PREDICT study (safety set) (reproduced from CS Table 57)

AEs	All Patients (n=736)	RAPID3-assigned (n=369)	CDAI-assigned (n=367)
[redacted])	[redacted]	[redacted]	[redacted]
Events, n (patient %) [#][a]			
All AEs	559 (76.0) [2,145]	270 (73.2) [1,070]	289 (78.7) [1,075]
Severe AEs	76 (10.3) [120]	37 (10.0) [65]	39 (10.6) [55]
SAEs	71 (9.6) [112]	32 (8.7) [50]	39 (10.6) [62]
Discontinuations due to AEs	78 (10.6) [112]	32 (8.7) [49]	46 (12.5) [63]
Drug-related AEs[b]	173 (23.5) [325]	88 (23.8) [156]	85 (23.2) [169]
AEs leading to death	2 (0.3) [6]	0	2 (0.5) [6]

AE: adverse event
[a] Number of individual AE occurrences; [b] AEs with relationship of 'related' or those with missing responses
Adapted from Curtis *et al.* (2015)[57]

Table 21: Safety up to the end of the 24 week double-blind J-RAPID study (overall safety population) (reproduced from company's clarification response to ERG)

Exposure and AEs	CZP 200 mg Q2W + MTX (n=82)	PBO + MTX (n=77)
Duration of exposure, PY	36.80	25.16
Any AEs by maximum intensity, n (%)	63 (76.8)	51 (66.2)
Mild	41 (50.0)	28 (36.4)
Moderate	20 (24.4)	22 (28.6)
Severe[a]	2 (2.4)	1 (1.3)
Treatment-related[b], n (%)	31 (37.8)	21 (27.3)
Serious AE (total), n (%)	4 (4.9)[c]	1 (1.3)
Malignancy, n (%)	0	0
Deaths, n (%)	0	0
Most common AEs[d] (≥5% in any group), n (%)		
Nasopharyngitis	11 (13.4)	9 (11.7)
Abnormal hepatic function	3 (3.7)	4 (5.2)
RA exacerbation	4 (4.9)	9 (11.7)
Pharyngitis	5 (6.1)	3 (3.9)
Serious AEs, n (%)		
RA	1 (1.2)	0
Bronchitis	1 (1.2)	0
Pyelonephritis	1 (1.2)	0
Purulent myositis	1 (1.2)	0
Subcutaneous tissue abscess	1 (1.2)	0
Urosepsis	1 (1.2)	0
Anal fistula	0	1 (1.3)

PY: patient years
[a] Severe AE defined as an event that prevents work or daily activities.
[b] Treatment-emergent AEs for which the relationship to the study drug cannot be ruled out.
[c] 6 events in 4 patients.
[d] Preferred terms according to MedDRA terminology.
Adapted from Yamamoto *et al.* (2014[32])

Table 22: Safety up to the end of the 24 week double-blind HIKARI study (overall safety population) (reproduced from company's clarification response to ERG)

Exposure and AEs	CZP 200 mg Q2W (n=116)	PBO (n=114)
Duration of exposure, PY	49.43	34.08
Any AEs by maximum intensity, n (%)	83 (71.6)	67 (58.8)
Mild	33 (28.4)	29 (25.4)
Moderate	44 (37.9)	36 (31.6)
Severe[a]	6 (5.2)	2 (1.8)
Treatment-related, n (%)	44 (37.9)	24 (21.1)
Serious AE (total), n (%)	13 (11.2)[b]	3 (2.6)[c]
Deaths, n (%)	1 (0.9)	0
AEs leading to withdrawal	9 (7.8)	3 (2.6)
Most common AEs[d] (≥3% in any group), n (%)		
Nasopharyngitis	20 (17.2)	16 (14.0)
Rash	10 (8.6)	0
Pharyngitis	6 (5.2)	5 (4.4)
Eczema	6 (5.2)	3 (2.6)
RA	5 (4.3)	14 (12.3)
Abnormal hepatic function	4 (3.4)	4 (3.5)
Hypertension	4 (3.4)	1 (0.9)
Constipation	4 (3.4)	0
Upper respiratory tract infection	3 (2.6)	4 (3.5)
Serious AEs, n (%)	4 (3.4)	1 (0.9)[e]

PY: patient years.
[a] Severe AE defined as an event that prevents work or daily activities.
[b] 14 events in 13 patients.
[c] 5 events in three patients.
[d] Preferred terms according to MedDRA terminology.
[e] 2 events in the same patient.
Adapted from Yamamoto *et al.* (2014[33])

Table 23: **Crude incidence (unadjusted) and adjusted risk (HR) of first hospitalised infection from Yun *et al.* 2016[51] (reproduced from CS Table 59)**

Drug	n	Events[a]	PYs	Crude IR/100 PYs (95% CI)	Crude absolute risk difference	Adjusted HRb (95% CI)
ABA	9204	705	5377	13.1 (12.2, 14.1)	Reference group	Reference group
ADA	4845	317	2171	14.6 (13.1, 16.3)	0.015	1.08 (0.93, 1.25)
CZP	1866	106	747	14.2 (11.7, 17.2)	0.011	1.07 (0.86, 1.32)
GOL	3814	87	616	14.1 (11.5, 17.4)	0.010	1.14 (0.90, 1.44)
ETA	1394	275	1726	15.9 (14.2, 17.9)	0.028	1.24 (1.07, 1.45)*
IFX	3944	370	2178	17.0 (15.3, 18.8)	0.039	1.39 (1.21, 1.60)*
RTX	4718	541	2898	18.7 (17.2, 20.3)	0.056	1.36 (1.21, 1.53)*
TOC	2016	129	863	14.9 (12.6, 17.8)	0.018	1.10 (0.89, 1.34)

*Considered significant by the authors

[a] First hospitalised infection during follow-up; [b] Adjusted for infection risk score, number of previous biologics used, disability status, glucocorticoid use at baseline, MTX use at baseline, most recent biologic prior to baseline and Medicaid eligibility; ABA was used as the reference group.

Table 24: Type of infection, number of hospitalised infections, and mortality associated with different biologic agents* (Yun *et al.* 2016[51])

Drug	ABA	ADA	CZP	ETA	GOL	IFX	RTX	TOC
Total no. of infections**	926	397	116	336	99	472	643	134
Septicemia/bacteremia	15.4	15.6	19.8	18.8	15.2	16.7	17.3	18.7
Pneumonia/upper respiratory tract infection	29.9	31.7	30.2	31.3	32.3	35.2	35.9	32.1
Skin and soft tissue infection	12.9	12.9	10.3	11.9	9.1	10.8	10.9	13.4
Genitourinary tract infection	28.8	26.5	29.3	26.2	35.4	24.4	21.8	22.4
Other	12.9	10.5	8.8	10.2	12.6	10.7	11.7	12.7
Length of hospital stay for serious infection, mean ±SD days	9.2611.3	8.9610.4	10.8613.8	10.6612.0	9.5617.8	11.1615.9	9.169.1	10.0613.1
Mortality during or within 30 days after hospitalisation	5.7	5.3	7.8	4.5	4.0	5.1	4.5	5.9

* Except where indicated otherwise, values are % ** The total number of infections is greater than the total number of outcome events shown in Table 2, as patients may experience multiple types of infection during a single hospitalisation

4.2.17 External validity

The CZP RCT with the largest TNFi-experienced population was REALISTIC (CZP n=320, PBO n=80), which included patients receiving CZP as monotherapy and also in combination with background cDMARDs. The clinical advisors to the ERG believed the REALISTIC trial was largely appropriate to the decision problem although it is commented that this trial excluded patients who had been treated with prior RTX.

There were low numbers of TNFi-experienced patients in some of the CZP RCTs (e.g. J-RAPID [CZP n=11, PBO n=15] and HIKARI [CZP n=6, PBO n=10]). The only CZP RCT that recruited solely TNFi-experienced patients was SWITCH, which also had a small sample size (CZP n= 27, PBO n= 10). J-RAPID and HIKARI were performed in Japan only. The mean dose of MTX at baseline was lower in J-RAPID (set to 6-8 mg/week in line with the licensed dose in Japan at the time of study). None of the included CZP RCTs recruited patients from UK centres. The CS included observational efficacy evidence from the ARTIS Swedish registry-based study.

A number of patient groups were excluded from participation in the CZP RCTs, for example, patients with inflammatory arthritis other than RA, a history of chronic or serious infection, any current infection, a history of or current TB, hepatitis B, hepatitis C, malignancy, lymphoproliferative disorder, demyelinating disease, and congestive heart failure. Patients with any such conditions would not be represented by the included trial populations.

REALISTIC, J-RAPID and HIKARI excluded patients who had received more than 2 TNFis. REALISTIC also excluded patients who had received RTX or ABA. Patients were not permitted to enter PREDICT if they had been treated with ≥ 3 TNFis or any non-TNFi biologic. Therefore, the included trial populations would not reflect patients with a history of extensive biologic treatment.

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

Appendix 8.12.3 of the CS provides network diagrams showing the evidence available for each outcome measure at different assessment times (i.e. ACR20 at 3 months, ACR50 at 3 months, EULAR (good/moderate) at 3 months, EULAR (good) at 3 months, ACR20 at 6 months, ACR50 at 6 months, ACR70 at 6 months, EULAR (good/moderate) at 6 months, EULAR (good) at 6 months. Analyses were performed separately and results were presented as odds ratios for six of these combinations of outcome measure and assessment time assuming a binomial likelihood. Different outcome measures at common assessment times (e.g. ACR20 and ACR50 at 3 months) could have been analysed in one model (i.e. a multinomial likelihood) if proportional odds could be assumed.

Appendix 8.12.5 of the CS provides an assessment of model fit. The company asserts that a fixed effect model is preferred because of the limited number of studies included in the NMAs. A fixed effect model answers the question, "Did the treatments have an effect in the studies included in the analysis?". Alternatively, if it can be assumed that all studies comparing specific pairs of treatments all estimate the same treatment effects (which would be plausible if they all followed the same protocol) then a fixed effect model can be assumed. If it cannot be assumed that studies comparing specific pairs of treatments are estimating the same treatment effects, then a random effects model should be assumed. A random effects model would also be appropriate if we expect heterogeneity and the question being answered is, "Will the treatments have an effect when given to future patients?", which is generally the more relevant question of interest.

Prior information should not be used unthinkingly in a Bayesian (network) meta-analysis. When there are insufficient sample data with which to update prior distributions, the prior information will be influential and not uninformative. In such situations, if the prior information does not represent reasonable prior beliefs, then the results will not represent reasonable posterior beliefs. Reference prior distributions have been used in the random effects analyses presented in the CS resulting in posterior estimates of the between-study standard deviations that are greater than one. Whilst heterogeneity is expected,[58] values for the between-study standard deviation greater than one imply extreme heterogeneity which the ERG consider to be implausible.

The ERG considers that the random effects model has been implemented inappropriately because it failed to recognise the implausible influence that the reference prior distribution for the between-study standard deviation has on the results, and that the fixed effect model answers a limited question. Results from the former will be inappropriately imprecise, whilst results from the latter will be inappropriately precise. The ERG would have preferred to see results from a random effects model incorporating weakly informative prior information for the between-study standard deviation reflecting plausible prior beliefs. In the ERG's clarification letter (Question A20), the company was asked to provide estimates and 95% credible intervals for the between-study standard deviations where these were not provided in the original CS. The company responded by stating that the between-study variances were provided in Appendix 8.12.5. However, these were only for the separate analyses of each ACR and EULAR category using a binomial likelihood. A random effects model was not used to model ACR and EULAR when using a multinomial likelihood function allowing for the different response categories on the probit scale; hence, the between-study standard deviation was assumed by the company to be zero.

The description of the model in Section 4.10.5.1 of the CS is not as would typically be implemented, although it may not have been implemented in the same way as it has been described. In addition, the

company asserts that the Bayesian "approach relies on vague prior distributions on study and treatment effects, and results in posterior distributions for relative and absolute effects." The purpose of a meta-analysis is to estimate relative treatment effects and, as conventionally implemented, it can only provide estimates of absolute treatment effects by combining it with an external estimate of a suitable baseline response for the reference treatment. The company specifically distinguishes between study effects and treatment effects and yet wrote that the study effects are assume to arise from a population of study effects; this would be a random baseline model which would have the effect of breaking the blind and is generally not recommended. A conventional random (treatment) effects meta-analysis assumes that it is the treatment effects from each study that are assumed to arise from a population of treatment effects. It is wrong, as with any Bayesian analysis, to suggest that the analysis relies on vague prior distributions; prior distributions should not be used unthinkingly, especially for variance parameters, because these are seldom non-informative, as is evident from the results of the random effects models presented in the CS.

The company presented an assessment of the relative goodness-of-fit of the fixed and random effects models using the deviance information criterion (DIC). However, DIC cannot tell us anything about the appropriateness of the prior distributions; in analyses with few studies, the DIC cannot tell us which of the models is the most appropriate and the DIC values are not helpful in this context. The company also provided an assessment of the absolute goodness-of-fit of the models using residual deviance but did not include in the tables the number of data points to which the models were fitted, making it impossible to assess absolute goodness-of-fit.

WinBUGS code for the fixed effect model for binary data is presented in Appendix 8.13.1 of the CS and includes the following statements:

```
for (i in 3:NT)
        {
        prd.m[i] <- pr.m[i]+d[2]-d[1]
        d[i]~dnorm(prd.m[i],pr.prec[i])
        }
```

The ERG is unable to identify the interpretation of the parameters pr.m[i] and prd.m[i]. In addition, it appears that a hierarchical random effect is being placed on the treatment effects relative to the reference treatment in spite of already giving them independent prior distributions.

WinBUGS code for the random effects model for binary data is presented in Appendix 8.13.2 of the CS and includes the following statements (as above):

```
for (i in 3:NT)
        {
        prd.m[i] <- pr.m[i]+d[2]-d[1]
        d[i]~dnorm(prd.m[i],pr.prec[i])
        }
```

Again, the ERG is unable to identify the interpretation of the parameters pr.m[i] and prd.m[i] and it appears that a hierarchical random effect is being placed on the treatment effects relative to the reference treatment in spite of already giving them independent prior distributions.

The ERG questions the interpretation of the results made by the company: wide credible intervals reflect uncertainty about the true treatment effect and not that there is minimal difference between treatments; the reference to frequentist significance levels is inappropriate in these Bayesian analyses; an inability to assert which is the more effective treatment based on the available evidence does not mean that the treatments are equally effective.

During the clarification process (see clarification response,[28] Question A28b) the company was asked to clarify why Bucher analyses were preferred to a Bayesian Random effects model. The company recognised that the prior distribution for the between-study standard deviation was not non-informative but did not consider using a plausible weakly informative prior distribution, and instead chose to assume that the between-study standard deviation was zero.

The company's clarification response (clarification question A30[28]) confirmed that the results presented in Section 4.10.6 of the CS are random effects means and not predictive distributions for the effects in new studies. The ERG considers the random effects means to underestimate the true uncertainty about the treatment effect in a heterogeneous population.

The company was also asked to clarify (clarification question A31) what is considered to be a clinically relevant non-inferiority margin for claims of CZP+MTX being "at least as effective" as other comparators. The company did not provide a specific clinically relevant equivalence margin.

The results from the NMA are shown in

Figure 13

Figure 13, Figure 14**, and**

.

Figure 13: Estimated mean EULAR response probabilities from the NMA for Population A (reproduced from Figure 58 of the CS)

Figure 14: Estimated mean EULAR response probabilities from the NMA for Population B (produced with company's model)

Figure 15: Estimated mean EULAR response probabilities from the NMA for Population C (reproduced from Figure 60 of the CS)

4.4 Additional work on clinical effectiveness undertaken by the ERG

No additional analyses were undertaken by the ERG for the reasons summarised in Section 4.5.

4.5 Conclusions

The clinical effectiveness systematic review included six CZP RCTs (REALISTIC, DOSEFLEX, PREDICT, SWITCH, J-RAPID and HIKARI). The ERG was satisfied that the clinical effectiveness searches were likely to have identified all relevant published RCT evidence. The ERG considered that the study selection eligibility criteria were generally consistent with the decision problem as outlined in the final NICE scope. The clinical advisors to the ERG did not highlight any additional relevant RCTs that should have been included in the CS. A CZP RCT by Kang *et al.* (2012) was identified by the ERG. The company clarified to the ERG that the Kang trial was not included in the CS because only low numbers of patients in the trial were TNFi-experienced. However, since two CZP RCTs were included in the CS that also had low numbers of TNFi-experienced patients (J-RAPID and HIKARI), the ERG considered that additional justification should have been provided by the company to support their decision to exclude the Kang trial. All six included CZP RCTs were considered to be of good quality. The company also included supplementary observational evidence from the Swedish registry-based study ARTIS. Safety evidence was summarised from the included CZP RCTs, a pooled analysis of CZP safety and three additional safety studies (Yun *et al.* 2016; Simard *et al.* 2011, Curtis *et al.* 2015).

The coverage of disease activity and physical function outcomes and safety in the CS was good, with data presented for ACR and EULAR responses, DAS28, CDAI, HAQ-DI and adverse events. More limited findings were reported for fatigue, pain and health-related quality of life. Data were not included in the CS for the following outcomes specified in the final NICE scope: radiological progression, joint damage or extra-articular manifestations of disease.

ACR response data were presented for all included CZP RCTs (PREDICT reported modified ACR (mACR). The CZP RCTs included in the CS [redacted] Findings from the classical meta-analysis indicated better ACR20, ACR50 and ACR70 responses at 3 months for CZP + MTX-treated patients compared with PBO. Data for ACR response in TNFi-experienced patients receiving CZP as monotherapy were available for two PBO-controlled CZP RCTs (REALISTIC and HIKARI). The evidence for ACR response in TNFi-experienced patients receiving CZP as monotherapy [redacted]. Results from the classical meta-analysis indicated a more favourable ACR response for CZP monotherapy patients compared with PBO, although the results were inconclusive.

EULAR response data were available for all included CZP RCTs. Classical meta-analyses of EULAR data at 3 months were performed to compare the effects of CZP (both in combination with MTX and as monotherapy) with PBO. The results indicated more favourable EULAR responses for the CZP intervention groups, although this effect appeared weaker for CZP as monotherapy and results were inconclusive.

Four RCTs (REALISTIC, DOSEFLEX, PREDICT and SWITCH) presented HAQ-DI scores (or MD-HAQ/M-HAQ for PREDICT). CZP-treated patients reported [redacted] Limited data were presented for the outcomes of pain and fatigue, with patients receiving CZP experiencing [redacted] Health-related quality of life data were presented in the CS for DOSEFLEX only, with the CZP group experiencing [redacted]

Data from the ARTIS registry study were presented in the CS, showing significant benefits of CZP treatment in TNFi-experienced patients in DAS28 ($p<0.0001$) and HAQ ($p<0.0001$) at 3 and 6 months following the start of CZP therapy.

Since no head-to-head RCTs were identified for the comparison of CZP with comparator bDMARDs, indirect treatment comparisons were performed. Nine RCTs of CZP and comparator bDMARDs were included in the indirect treatment comparisons.

The ERG believes that there were several limitations with the NMA as presented within the CS. Several changes are required to the analyses and reporting of the results in order for them to represent genuine uncertainty and useful for decision-making purposes, including: incorporating weakly informative prior information for the between-study standard deviation; generating predictive distributions of the effects of treatments in a new study; using the evidence from the REALISTIC study to generate the probabilities of being in each ACR and EULAR category for the reference treatment; and taking draws from the joint posterior distribution of treatment effects rather than assuming univariate normal distributions for them. It was not possible for the ERG in the time available to make the required changes to produce robust results and the ERG has not amended the NMA presented in the CS.

5 COST EFFECTIVENESS

5.1 Comment on company's review of cost-effectiveness evidence

5.1.1 The objective of cost effectiveness review

The company conducted a systematic literature review to identify relevant published economic evaluations and cost & resource use studies of patients with RA who had previously been exposed to TNF inhibitors.

Searches were run on Embase, MEDLINE and the Cochrane Library (HTAs and economic evaluations) and were restricted to evidence published since 2005 due to the identification of a previous MTA (TA195[25]) covering studies up to that date. The CS used inclusion criteria based on this review, but expanded the list of interventions of interest to include CZP, GOL and TOC.

Unfortunately, as with the clinical effectiveness search, the company searched Embase and Medline simultaneously meaning that separate results from each are not reported. This also led the ERG to question the source of the filter used to identify economic and cost-resource use studies, since most published filters are optimised for use only on one specific database.

In response to a request for clarification from the ERG (see clarification response,[28] question A4), the company stated that filters were based on those developed by SIGN.[59] It should be noted that SIGN filters are published unvalidated, and that the company have made numerous alterations to the economic studies filter. Furthermore, although the inclusion criteria (CS, Table 60, p.163) state a range of study types of interest (including quality of life studies as well as cost-benefit analyses) the search terms used in the filter are primarily drawn from SIGN's economic filters. However, the company have drawn on data from previous NICE technology appraisals to supplement the published literature.

Unusually for an STA submission, the CS does not include an original review of HRQoL evidence, but uses data from the trials identified in the clinical effectiveness review and from a review of mapping algorithms in RA by Pennington & Davis.[60] The ERG queried this in the clarification letter, since after consulting the Pennington & Davis review it became apparent that their review was based on a rudimentary search of one database (Medline). In response to the request for clarification from the ERG (see clarification response,[28] question A3) the company stated that the that the HERC database[61] had been searched to confirm that there had been no new mapping studies since Pennington & Davis conducted their review.

Despite these issues, the ERG considers the company's searches for existing economic evaluations to be adequate and that it is likely that all relevant studies have been identified.

5.1.2 Inclusion and exclusion criteria used in study selection

The systematic literature review conducted by the company to identify cost-effectiveness studies relevant to the decision problem used the inclusion and exclusion criteria listed in Table 25.

Table 25: Eligibility criteria used in the study selection

Dimension	Inclusion criteria	Exclusion criteria
Disease	Moderate to severe RA	Disease activity (moderate to severe) unclear
Population	Adults	Children
Line of therapy	Patients who have had an inadequate response to a TNFi	Patients not previously treated with any TNFi or not controlled to these drugs
Interventions	• Adalimumab (ADA) • Etanercept (ETA) • Infliximab (IFX) • Rituximab (RTX) • Abatacept (ABA) • Certolizumab pegol (CZP) • Tocilizumab (TOC) • Golimumab (GOL)	No specific intervention present in the study
Outcomes	• Cost effectiveness/utility • Cost estimates • Quality of life estimates	
Study design	• Cost-consequence/benefit analyses • Cost effectiveness/utility analyses • UK-based cost studies • Cost-of-illness studies • Quality of life studies	

RA=rheumatoid arthritis; TNFi = tumour necrosis factors inhibitor; UK = United Kingdom

5.1.3 Findings and conclusions of the cost effectiveness review

The systematic review undertaken by the company identified 3,861 unique records. Of these records, 3,784 records were excluded based on their title or abstract. Of the remaining 77 publications, 48 were excluded after full text screening. The most frequent reason for exclusion (36 out of 48) after full text screening was "wrong line of therapy". The remaining 29 publications reported the results of 23 economic evaluations. Two of these evaluations were excluded because they were budget impact analyses.

The company performed an additional *ad hoc* search of previous NICE TAs in the NICE website. This search identified two STAs (TA225 and TA247) and one MTA (TA195) that reported on the

cost-effectiveness of drug therapies in patients who had failed on a previous TNFi. These evaluations were included in the review alongside the published evaluations.

The CS does not present any conclusions from the cost-effectiveness review; instead, the CS argues that none of the identified studies captured the cost-effectiveness of CZP in patients who had failed on a previous TNFi. As such, the company presented the cost-effectiveness results from a *de novo* model developed for this appraisal (see Section 5.2). The CS does not clarify whether the models identified during the review were used to inform the structure or the parameters of the *de novo* model.

5.2 Summary of the company's economic evaluation

This section summarises the company's economic evaluation. The ERG's critical appraisal of the company's economic evaluation is described in Section 5.3. The ERG notes that the company submitted a revised model after the clarification process; unless otherwise stated, the revised model is described in this chapter.

5.2.1 NICE Reference Case checklist

A summary of the key features of the company's revised *de novo* model after the clarification process is provided in Table 26.

Table 26: Key features of the company's revised *de novo* model after the clarification process

Feature	Parameter value
Population, intervention, comparators and outcomes	See Sections 3.1 to 3.4
Model Structure	Markov cohort model
Starting age (years)	[redacted]
Time horizon	45 years, assumed representative of lifetime
Cycle length	6 months
Half-cycle correction	Included
Measure of health effects	QALYs
Primary health economic outcome	Incremental cost per QALY gained
Discounting	Costs and benefits were discounted at 3.5% per annum
Perspective	The NHS and Personal Social Services (PSS) in England
Price year	2015

5.2.2 Population

The target population of the CS is defined as patients with active moderate to severe RA whose disease has not responded adequately to a TNFi. Moderate to severe RA is defined as a DAS28(ESR) score greater than 3.2. The company defined three subpopulations for the cost-effectiveness analysis based on contraindications to, or withdrawals due to AEs of, RTX and/or MTX:

A. Adults previously treated with other DMARDs including at least one TNFi
B. Adults for whom RTX is contraindicated or withdrawn
C. Adults for whom RTX therapy cannot be given because MTX is contraindicated or withdrawn.

The ERG notes that the company failed to specify that only those patients eligible for RTX + MTX were analysed within Population A.

The company claims that due to limited data the fourth population mentioned in the scope, 'people with moderate to severe active disease despite treatment with biological DMARDs recommended according to NICE guidance', which had a comparator of best supportive care was not included in the company's health economic analysis. From the previous experience of the ERG in modelling RA, and from inferences that can be made from results presented in the CS, it is expected that the cost-effectiveness of CZP would be worse compared with best supportive care than other bDMARDs. However, the ERG acknowledge that the current standard of care is bDMARD, and that this would not change based on the recommendation of this STA. As such, the ERG believes that the comparison against bDMARDs is appropriate.

Baseline characteristics of the model population were based on mean estimates from the TNF-IR patients of the REALISTIC trial. The same baseline characteristics were used for all three subpopulations due to insufficient data to differentiate between them. The average age of the cohort was [redacted] years, with [redacted] being female and a mean baseline HAQ score of [redacted]. Clinical advisors to the ERG suggested that these values appear representative of the target population.

5.2.3 Interventions and comparators

Interventions and comparators differ depending on the subpopulation. For Population A, the company defined the intervention and comparator sequences shown in Table 27. The intervention sequence only differs from the comparator sequence in that it includes an extra line of therapy, CZP + MTX, at the beginning of the sequence. The ERG notes that the intervention sequence includes the treatment defined as the comparator in the scope (RTX + MTX) and therefore represents a comparison of an elongated sequence compared with a standard sequence. Other potential sequences, such as replacing RTX + MTX with CZP + MTX, or comparing the elongated sequence with an equally long sequence with RTX + MTX before CZP + MTX were not considered within the CS. The ERG therefore believe that a fully incremental analysis of all appropriate sequences has not be undertaken.

Furthermore, ABA + MTX is included after TOC + MTX in the intervention and comparator sequences despite ABA + MTX not being recommended by NICE after the failure of a TNFi where

patients are eligible for RTX + MTX. The clinical advisors to the ERG indicate that four lines of bDMARDs does not represent clinical practice.

Table 27: Intervention and comparator sequences in patients eligible for RTX + MTX (Population A)

Line of therapy	Intervention	Comparator
First	CZP + MTX	RTX + MTX
Second	RTX + MTX	TOC + MTX
Third	TOC + MTX	ABA + MTX
Fourth	ABA + MTX	MTX + HCQ + SSZ
Fifth	MTX + HCQ + SSZ	NBT†
Sixth	NBT†	Palliative care
Seventh	Palliative care	-

†Non-biologic treatment: a weighted mix of leflunomide, gold, ciclosporin, azathioprine (25% each)

The intervention and comparator sequences analysed for Population B, that is, patients for whom RTX is contraindicated or has been withdrawn due to an AE, are shown in Table 28. Only the first line of therapy differs between the intervention and comparator sequences. Unlike the analyses conducted for Population A, non-biologic treatments (LEF, GLD, CIC and AZA) were considered separate lines of therapy for Population B.

Table 28: Intervention and comparator sequences in patients for whom RTX is contraindicated or withdrawn (Population B)

Line of therapy	Intervention	Comparators
First	CZP + MTX	Comparator biologic† + MTX
Second	MTX + HCQ + SSZ	MTX + HCQ + SSZ
Third	Leflunomide	Leflunomide
Fourth	Gold injection	Gold injection
Fifth	Ciclosporin	Ciclosporin
Sixth	Azathioprine	Azathioprine
Seventh	Palliative care	Palliative care

† ABA, ADA, ETA, GOL, IFX or TOC

The intervention and comparator sequences analysed for Population C, that is, patients for whom RTX therapy cannot be given because MTX is contraindicated or withdrawn, are shown in Table 29. Only the first line of therapy differs between the intervention and comparator sequences: the first treatment in the sequence is CZP monotherapy in the intervention sequence and ADA, ETA or TOC in the comparator sequences. As with Population B, non-biologics are considered sequentially as different lines of treatment for Population C.

Table 29: Intervention and comparator sequences in patients for whom MTX is contraindicated or withdrawn (Population C)

Line of therapy	Intervention	Comparators
First	CZP	Comparator biologic†
Second	Leflunomide	Leflunomide
Third	Gold injection	Gold injection
Fourth	Ciclosporin	Ciclosporin
Fifth	Azathioprine	Azathioprine
Sixth	Palliative care	Palliative care

† ADA, ETA or TOC

5.2.4 Perspective, time horizon and discounting

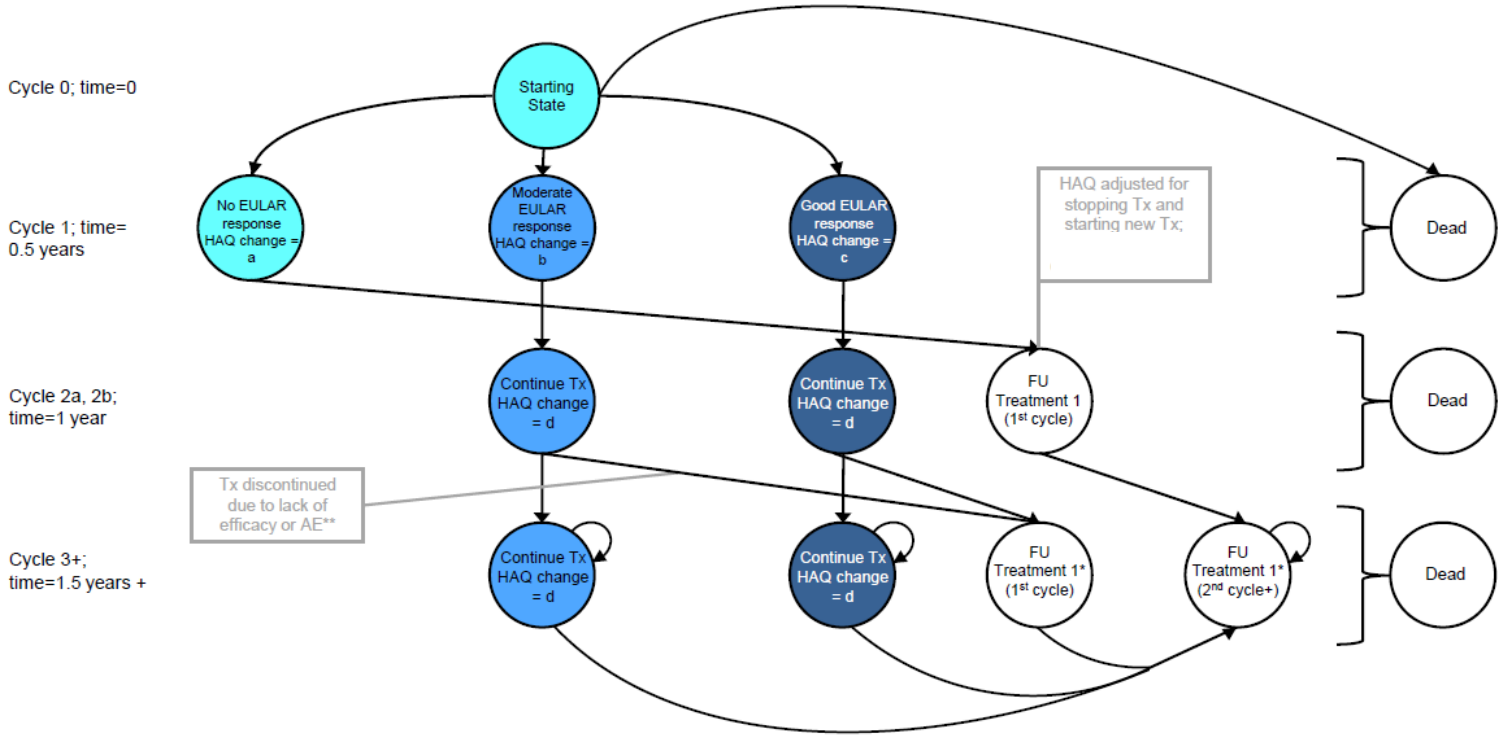
Base case costs and health outcomes are evaluated from the perspective of the NHS in England and PSS. The time horizon used in the model is 45 years which, considering a mean starting age of [redacted] years, is considered representative of a lifetime horizon. In accordance with the NICE Reference Case, the model includes discounting of both costs and effects at an annual rate of 3.5%.[62]

5.2.5 Model structure

The model provided by the company is a Markov cohort model constructed in Microsoft Excel©. The model contains 19 states: the starting state; three states to represent the different EULAR responses (no response, moderate or good) to the first treatment; two states for each of the seven subsequent treatments (one state to represent the first six months of treatment, and one state for the remainder of time on that treatment); and death. Although the model has the capacity to analyse eight lines of treatment, the strategies evaluated within the CS include only a maximum of seven lines of therapy.

Patients enter the model in the starting state where the first treatment is initiated. Patients then transit to one of three states representing their EULAR response (no response, moderate or good) to the first treatment. Patients who experience no EULAR response automatically transition to the first state for the next treatment in the sequence, whereas the patients with good or moderate response stay in their respective states until treatment discontinuation. On receiving the next line of treatment, patients spend a cycle in a tunnel state that represents the first six months of the subsequent treatment. Following this, patients then transit either to a state representing the remainder of duration on that treatment, if a moderate or good EULAR response was observed, or to the next treatment in the sequence following no response. A schematic of the company's model is presented in Figure 16.

Figure 16: Transition diagram of the company's model (reproduced from Figure 57 of the CS)[31]



*Follow-up treatment states: duplicated for each follow-up treatment. Patients not responding in first 6 months of follow-up treatment will move to the next treatment in the sequence; **Reason for discontinuation (lack of efficacy) governed by probabilities after leaving treatment health state.
HAQ-DI categories relate to the non-treatment specific costs associated with disability.

5.2.6 Treatment effectiveness and extrapolation

Within the health economic model, treatment effects were modelled through treatment-dependent probabilities of EULAR responses. Response to treatment entails a change in HAQ score, which in turn affects the estimated HRQoL, mortality and costs.

5.2.6.1 EULAR response probabilities for first therapy

Response to the first biologic treatment was modelled through treatment-dependent probabilities for moderate and good EULAR responders. According to the CS, these probabilities were estimated using summary statistics on the trial-specific baseline effects, effect size estimates and cut-off statistics extracted from the NMA described in Section 4.1.3. The statistical model used in the NMA was a multinomial likelihood (implemented as a conditional binomial likelihood) with a probit link.

In the CS, the company stated that the equations for estimating the probability of response are applied as follows:

$$P\ (EULAR\ moderate\ ) = \Phi(\mu + \beta + \mathrm{Z}) - \Phi(\mu + \beta)$$

$$P\ (EULAR\ good) = 1 - \Phi(\mu + \beta + \mathrm{Z})$$

$$P\ (no\ response) = 1 - \mathrm{P(EULAR\ moderate)} - P(EULAR\ good),$$

such that μ is the trial-specific baseline effects, β is the effect size estimate for treatment compared with baseline, Z is the cut-off statistic and Φ is the inverse of the standard normal cumulative distribution.

The company made the following assumptions when modelling the efficacy of CZP and comparators in the economic model to overcome the scarcity of data on the efficacy of biologics after the failure of a TNFi:

- Placebo response probabilities for the comparison with CZP + MTX were derived mapping the response observed at week 12 in REALISTIC to response at six months via a mapping matrix generated from patient-level data collected in the RAPID 1 and 2 trials[63, 64] (see Table 30).
- The company's systematic review failed to identify studies reporting the efficacy of ADA, ETA and IFX in combination with MTX in TNFi-IR patients. In the absence of data, the company assumed their efficacy to be equivalent to that of GOL.
- Biosimilars to IFX were assumed to have the same efficacy as IFX.
- The efficacy of TOC and GOL monotherapies was calculated assuming that the relative efficacies of TOC + MTX and GOL + MTX compared to CZP + MTX were maintained when MTX was removed allowing the calculation of efficacy for TOC and GOL

monotherapy from the efficacy of CZP monotherapy. The efficacies of ADA and ETA monotherapies were assumed to be equal to GOL monotherapy.

The probabilities of EULAR response at six months estimated from the Bayesian NMA are presented in

Figure 13, Figure 14 **and**

for Populations A, B and C, respectively.

Table 30: Mapping matrix for EULAR response from 3-months to 6-months for placebo (N=326; in RAPID 1 and 2 pooled data set)

EULAR	No response at 6-months	EULAR moderate at 6-months	EULAR good at 6-months	Total
No response at 3-months	[redacted]	[redacted]	[redacted]	[redacted]
EULAR moderate at 3-months	[redacted]	[redacted]	[redacted]	[redacted]
EULAR good at 3-months	[redacted]	[redacted]	[redacted]	[redacted]

Intention to treat population. Missing data imputed using last observation carried forward

5.2.6.2 Progression of HAQ and pain during first therapy

The company's model assumes an improvement of the HAQ score during the first six months of biologic therapy, a time span known as the "response period". The mean improvement in HAQ score was estimated for each EULAR response using a linear regression model fitted to patient-level data from the REALISTIC study at week 28.

Table 31 shows the parameters for the linear regression. The ERG notes that the fact that the variable "Concomitant use of MTX" has a [redacted] indicates that the use of concomitant use of MTX leads to higher HAQ scores compared with concomitant use of other cDMARDs or monotherapy.

Table 31: Variables of the linear regression to estimate change in HAQ score from baseline

Parameter	Mean estimate	Standard error	95% CI	*p*-value
Intercept	[redacted]	[redacted]	[redacted]	[redacted]
Moderate response†	[redacted]	[redacted]	[redacted]	[redacted]

Good response†	[redacted]	[redacted]	[redacted]	[redacted]
Baseline HAQ score	[redacted]	[redacted]	[redacted]	[redacted]
Baseline pain score on VAS	[redacted]	[redacted]	[redacted]	[redacted]
Concomitant use of MTX‡	[redacted]	[redacted]	[redacted]	[redacted]

†Reference category in the regression is 'no response'
‡Versus other cDMARDs or monotherapy
VAS = Visual Analogue Scale (0 to 100)

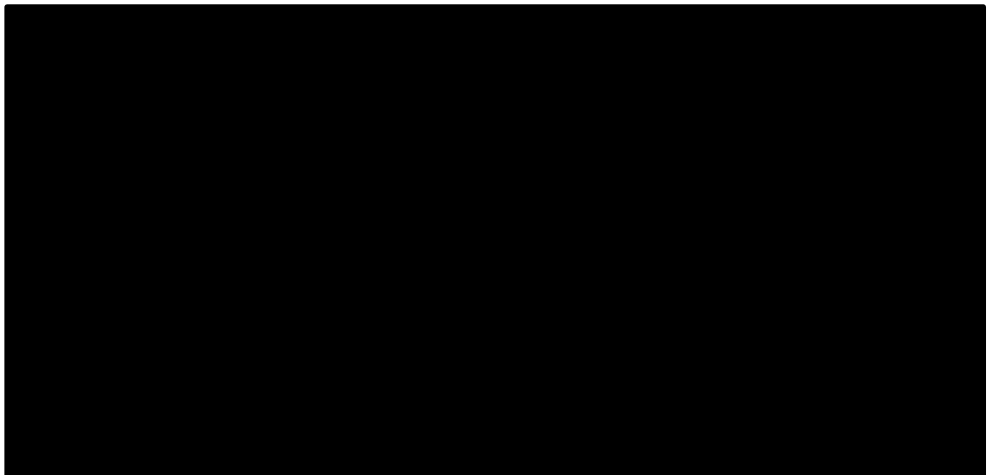
For baseline characteristics of the model population, the predicted changes in HAQ scores for non-responders, moderate responders and good responders were [redacted] for patients receiving concomitant use of MTX. These values were [redacted] for patients without concomitant use of MTX respectively. The ERG notes that these values are [redacted] than those used by the assessment group within TA375,[24] which were [redacted] for non-responders, moderate responders and good responders respectively. A scenario analysis undertaken by the company showed that using the values from TA375 made [redacted] As such, the ERG retained the values used by the company.

After the first six months of therapy, non-responders are assumed to discontinue the first treatment and start with the next treatment in the sequence. Moderate and good responders remain on treatment and their HAQ scores are assumed to remain constant. This assumption was made also by the Assessment Group within their independent economic analysis undertaken to inform TA375[24] based on data from the British Society for Rheumatology Biologics Register (BSRBR).

Pain scores, which are used along HAQ scores to estimate EQ-5D scores as described in Section 5.2.7, are calculated from HAQ scores fitting a linear regression to 100,398 from Data Bank. The fitted function is plotted against the data in

. A similar approach was used by the Assessment Group in TA375,[24] but they fitted a quadratic function to account for the drop in pain scores reported by patients with the highest HAQ scores.

Figure 17: Linear regression fitted to pain versus HAQ scores data from the US National Data Bank (reproduced from Figure 64 of the CS)



5.2.6.3 Discontinuation of first therapy

Time to discontinuation of first therapy was assumed to be equal for all biologics. This was modelled using a Weibull distribution; this approach is consistent with the Assessment Group's model in TA195.[25] The Weibull survival curve was fitted to time to discontinuation data of second-line TNFis taken from the BSRBR. The resulting parameters of the Weibull distribution for TNFis were scale=0.4416 (SE=0.00958) and shape=0.7008 (SE=0.03368). However, unlike the Assessment Group's model in TA195, which reported an average duration on RTX of 11.31 years, and 6.17 years on ABA, compared with 4.06 for TNFis, the company's model assumed the duration of treatment for TNFis could be applied to RTX and ABA. The duration on TOC was also assumed to equal that for the TNFis. The transition probabilities for therapy discontinuation for each cycle were appropriately calculated from the cumulative survival function.

Patients discontinuing treatment are assumed to experience a rebound and are therefore assigned an increase in HAQ score equal to that applied for the initial response to treatment. However, the benefits of the follow up treatment are immediately applied to the HAQ score.

5.2.6.4 Efficacy and discontinuation of follow up therapies

The efficacy of follow up treatments was modelled both in terms of treatment-dependent probabilities of EULAR response and treatment-dependent improvements in HAQ score for EULAR responders. Both the probabilities of response and the levels of HAQ improvement were estimated based on the RADIATE study,[40] which analysed the efficacy of TOC + MTX compared with PBO + MTX in patients who had failed to respond to one or more TNFis. The choice of this source to estimate the efficacy of follow up treatments was justified by the company due to the absence of data on patients who had received two or more TNFis and the fact that approximately 50% of the patients in RADIATE had received two or more TNFis. The efficacy of TOC + MTX was deemed to be applicable to all subsequent bDMARD treatments whilst the efficacy of PBO + MTX was deemed to

be applicable to all cDMARDs. The HAQ score improvement applied to EULAR responders was -0.39 for patients on bDMARDs and -0.05 for patients on cDMARDs.

Following the first six months of therapy that produced a good or moderate EULAR response, the HAQ score was assumed to remain constant for subsequent bDMARDs. For cDMARDs and palliative care, the CS assumed that HAQ scores increased linearly at a rate per annum of 0.045 for cDMARDs and 0.06 for palliative care based on assumptions within previous NICE appraisals. The ERG notes that in the most recent NICE appraisal, TA375,[24] a non-linear approach was used which better reflected disease progression whilst on cDMARDs. To compensate for the non-linearity of the HAQ score progression, the company set the maximum mean HAQ score to 2.76. This threshold was based on the mean HAQ score reported by Hernández Alava *et al.*[65] for the group of patients of the US National Data Bank for Rheumatic Diseases (a patient group with an average RA duration of 31 years).

Patients who fail to achieve either a moderate or a good EULAR response after the first six months of the subsequent treatment are assumed to automatically discontinue the treatment (unless the patients are in palliative care) and immediately start receiving the next therapy in the sequence. For patients who achieve a moderate or good response, treatment discontinuation was modelled differently depending on whether they were being treated with bDMARDs or cDMARDs. For bDMARDs, the discontinuation rate was assumed to be constant and equal to the discontinuation rate between six months and one year according to the Weibull distribution (described in Section 5.2.6.3) used to model time to discontinuation for the first bDMARD therapy (15.6%). In contrast, discontinuation of subsequent cDMARD therapies was based on the percentage of DMARD users remaining on treatment after 1 and 5 years reported in Edwards *et al.*[66] an observational study on the use of cDMARDs in the UK. Assuming constant discontinuation rates, the company calculated six-monthly discontinuation probabilities for different cDMARDs. The resulting transition probabilities are shown in Table 32.

Table 32: Percentage of DMARD users remaining on treatment as reported by Edwards *et al.*[66] and derived six-monthly probabilities

Treatment	Percentage of DMARD users remaining on treatment (Edwards *et al.*[66])		Derived six-monthly probabilities
	At 1-year	At 5-years	
bDMARDs*	-	-	15.6%
MTX	78.00%	57.10%	3.8%
Gold injection	45.90%	17.60%	11.3%
Ciclosporin	62.00%	34.20%	7.2%
Azathioprine	56.90%	34.80%	6.0%

*Not in Edwards *et al*., added for reference

5.2.6.5 Mortality

The company calculated age-adjusted gender-specific annual probabilities of death from the latest interim life tables published by the Office for National Statistics (ONS).[67] In order to estimate general mortality accurately it also took into account the varying proportion of females at different ages.

The excess mortality associated with RA was modelled based on the HAQ score, with an increase in the HAQ score being associated with an increased mortality. The company's model considered that mortality was linked to the patient's HAQ score during each cycle, rather than of the baseline HAQ score. The CS used a mortality hazard ratio of 1.43 for each HAQ score point based on a paper by Norton *et al.*,[68] which reports that HAQ score assessed at 1 year was a significant predictor of mortality. Each unit of HAQ increase was associated with a mortality hazard ratio of 1.43 (95% CI 1.17 to 1.75) after adjustment for differences in clinical and demographic factors. In contrast, the Assessment Group's model in TA375[24] considered, based on a literature review by Michaud *et al.*,[69] that only baseline HAQ score was predictive of mortality since there was no evidence to assume that change in HAQ score had an impact in mortality.

5.2.7 Health-related quality of life

The company's model used evidence from two different sources to estimate HRQoL values for use in the model. EQ-5D data from the PREDICT study were used to calculate the baseline utility and to estimate the utility values for the first therapy conditional on the patient's EULAR response. For subsequent therapies, mapping algorithms were used to estimate the changes in EQ-5D scores based on the estimated changes in HAQ scores.

Baseline utility and utilities for the first therapy

The EQ-5D utilities reported at baseline by patients with prior TNFi use participating in the PREDICT study were assumed to be representative of baseline utility for the target population reflected in the company's model. The resulting baseline utility was [redacted].

In order to estimate the change in EQ-5D conditional on EULAR response to the first therapy, the company undertook a series of linear regression analyses with patient-level data from the PREDICT study. First, the company conducted a set of univariate analyses to identify the independent variables that showed a significant (p <0.05) association with change since baseline in EQ-5D. Age, gender, number of prior TNFi therapies and disease duration in years were rejected because their p-value was higher than 0.05) and only baseline EQ-5D utility was retained (p <0.0001). Then, starting from the set of variables that were not rejected in the previous step, the company used a backward stepwise

routine to remove the variables that did not contribute to the predictive power of the model, as measured by the Akaike Information Criterion (AIC). The resulting linear regression model, which included the baseline EQ-5D variable, produced the results shown in Table 33.

Table 33: Results of the linear regression to estimate change in EQ-5D since baseline

Parameter	Mean estimate	Standard error	95% CI lower bound	95% CI upper bound	*p*-value
Intercept	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Baseline EQ-5D	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Moderate response†	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Good response†	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

†Reference category in the regression is 'no response'

The results of the regression in combination with the baseline utility imply that the utility of non-responders increases by 0.117 [redacted], that of moderate responders by 0.262 [redacted], and that of good responders by 0.367 [redacted]. As with previous RA models, including that for TA375,[24] the change in utility is removed the moment the patient discontinues their first treatment and utilities rebound back to baseline.

Mapping from changes in HAQ to changes in EQ-5D

The company claims to have obtained the relevant quality of life data from a review of mapping algorithms in RA that was reported by Pennington and Davis.[60] However, none of the studies reviewed by Pennington and Davis were used to inform the company's model. On the contrary, the base case analysis uses the same simple linear mapping used by the company within its submission for TA375[24] which applies an increase of 0.2102 in EQ-5D per unit decrease in HAQ, a relationship which the company attributes to Brennan *et al.*[70] However, the ERG has been unable to verify this source with the description of the method used by Brennan *et al.*[70] described as a "mapping which imputes the EQ5D from all 42 components of the HAQ disability questionnaire data". The CS also includes scenario analyses where the mapping algorithm for the different classes within the mixture model of Hernandez Álava *et al.*[65] are applied separately. Concerns relating to this alternative mapping analysis are described in Section 5.3.

Adverse events

The company did not consider the effects of AEs on health utility claiming that there is no significant difference in the risk of AEs between the intervention and its comparators.

Utilities used in the cost-effectiveness analysis

Utility values used in the different states of the model are summarised in Table 34.

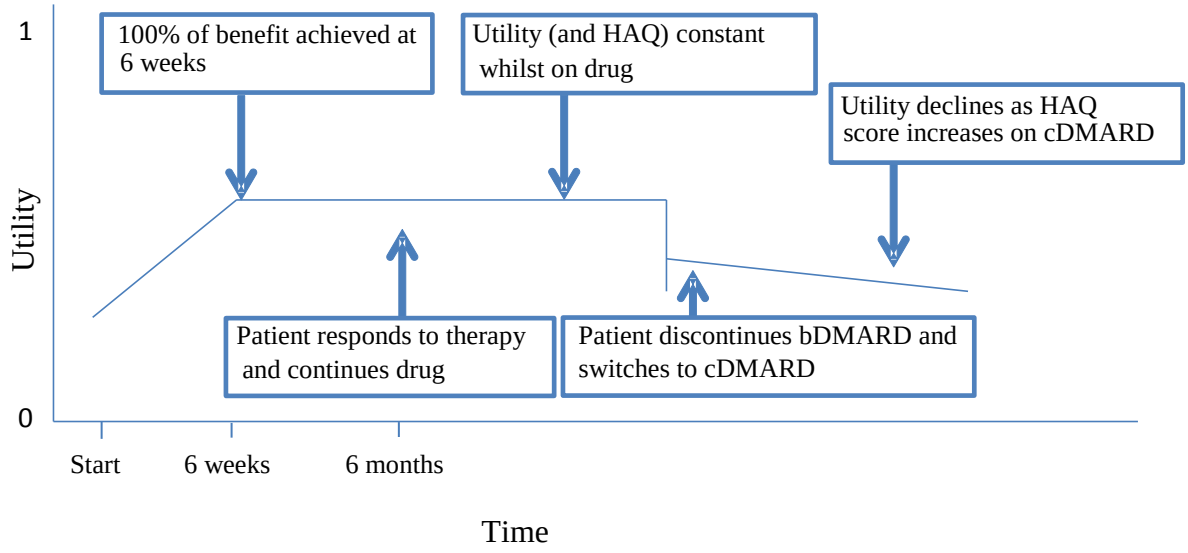
Table 34: Summary of utility values used in the base case analysis

Treatment	Percentage of DMARD users remaining on treatment (Edwards *et al*.[66])		Derived six-monthly probabilities
	At 1-year	At 5-years	
bDMARDs*	-	-	15.6%
MTX	78.00%	57.10%	3.8%
Gold injection	45.90%	17.60%	11.3%
Ciclosporin	62.00%	34.20%	7.2%
Azathioprine	56.90%	34.80%	6.0%

*Not in Edwards *et al*., added for reference

The patient population is assigned the baseline utility score of 0.4012 upon entry into the model. Conditional on their EULAR response, patients receive an increase of utility of 0.117 (no response), 0.262 (moderate response) or 0.367 (good response) with the base case analysis assuming a linear progression over a 6-week period to the final utility value. The utility of patients remains constant until they discontinue the first therapy. Patients then experience a 'rebound', suffering a decrease in utility equal to the improvement achieved in response to the first therapy. However, patients start the subsequent treatment immediately and its benefits are also immediately applied to the utility values. Patients discontinuing subsequent treatments also experience a rebound effect. Figure 18 illustrates an example of the progression of utility throughout the first bDMARD therapy followed by a cDMARD therapy: the utility remains constant during the bDMARD therapy but declines during the cDMARD therapy as HAQ score increases.

Figure 18: Illustration of an example of utility progression over time



5.2.8 Resources and costs

The company's economic analysis includes drug acquisition and administration costs, monitoring costs and additional resource use costs associated to the disease.

5.2.8.1 Drug acquisition costs

The drug acquisition costs used in the company's economic analysis were calculated from the recommended dosing schedules and the unit costs reported in the British National Formulary (BNF) 64.[72] The ERG confirmed that the unit costs reflect those listed in the most recently published version of the BNF.[30] Unit costs, unit doses and dosing regimens for bDMARDs and cDMARDs are listed in Table 35 and

Table **36**, respectively.

Four interventions have been recommended by NICE conditional on a Patient Access Scheme (PAS) being in place. For two of the interventions the PAS is public, these are: CZP in which treatment for patients during the first 12 weeks is provided to the NHS free of charge and GOL in which the company provides the 100 mg dose at the same price as the 50 mg dose. For two interventions, ABA and TOC, the PAS is confidential. The results presented in the CS include the PAS for CZP and GOL but not those for ABA or TOC.

For drugs in which the administration dose depends on the patient's body weight, an approximation of the weight distribution of patients in the REALISTIC trial was applied to calculate the average number of vials used. The base case assumes vial wastage, that is, that leftover drug from opened vials, is discarded.

Table 35: Unit costs, unit doses and dosing regimens of bDMARD used in the model

Treatment	Brand	Unit cost	Unit dose (mg)	Dosing regimen	Admins. / 6-month*	Cost first 6 months**	Annual cost (rest of treatment)**
CZP	Cimzia	£357.50	200	400 mg at week 0,2,4, and thereafter 200 mg every 2 weeks	16†/13	£2,145	£9,326
ABA (IV)	Orencia	£302.40	250	500 mg if < 60 kg, 1000 mg if > 60 kg, else 750 mg, weeks 0,2,4 thereafter every 4 weeks	8/6.5	£8,799	£14,330
ABA (SC)		£302.40	125	125 mg once per week	26	£7,862	£15,756
ADA	Humira	£352.14	40	40 mg every other week	13	£4,578	£9,187
ETA	Enbrel	£89.38	25	50 mg every week	26	£4,648	£9,327
	Benepali	£164	50			£4,264	£8,559
GOL	Simponi	£762.97	50	50 mg every month	6	£4,578	£9,187
IFX	Remicade	£419.62	100	3 mg/kg week 0, 2 and 6 thereafter every 8 weeks	5/3.50	£7,056	£9,910
	Inflectra / Remsima	£377.66				£6,438	£9,044
TOC (IV)	RoActemra	£256.00	200	8 mg/kg every 4 weeks	6.5	£7,553[#]	£15,138[#]
TOC (SC)		£228.28	162	162 mg once per week	26	£5,935	£11,902
RTX	MabThera	£873.15	500	Two doses of 1000 mg not more frequently than every 6 months	2‡	£3,840	£7,711[↕]

*For drugs with loading doses the number of doses for the first 6 months and subsequent 6-month periods are specified.
** Including drug acquisition and administration costs but not monitoring
† CZP patients are administered 16 doses in the first six months, but 10 are provided for free under the PAS
[#] Based on an average number of doses of 3.86 estimated by the company. The ERG estimated the average number of doses to be 3.44 as explained in Section 5.3.2, which results in a cost of £6,858 for the first six months and thereafter an annual cost of £13,748.
‡ A course of RTX consists of two doses of 1000 mg provided two weeks apart of each other.
[↕] Based on a retreatment interval of 6 months. The ERG preferred a retreatment interval of 7.35 months, as explained in Section 5.3.2 resulting in an annual cost of £6,300

Table 36: Unit costs, unit doses and dosing regimens of cDMARD used in the model

Drug acquisition	Unit cost (2016)	Unit dose (mg)	Units / administration	Administrations / 6-month period
MTX	£0.10	2.5	6	26
Hydroxychloroquine	£0.09	200	2	182.63
Sulfasalazine	£0.11	500	5	182.63
Leflunomide	£0.34	10	2	182.63
Gold injection	£4.56	10	1	26
Ciclosporin	£0.85	50	9	182.63
Azathioprine	£0.12	25	9	182.63
Prednisolone	£6.87	25	1	1.5

Loading doses were included for CZP, ABA (IV) and IFX as specified in Table 35.

5.2.8.2 Drug administration and monitoring costs

The company took into account the costs of administration and monitoring of drugs including the costs associated with outpatient visits for intravenous infusions, GP visits, monitoring tests and examinations. The costs of monitoring were calculated based on the unit costs listed in

Table 37 and the monitoring schedule shown in Table 38. These schedules were estimated from British Society of Rheumatology Guidelines for cDMARD therapies.[73] Given that this report does not include the bDMARDs, patients on biologic therapies were assumed to undergo the same monitoring schedule as that recommended for MTX. The company claims that patients on TOC would require additional monitoring for neutrophils, platelets and lipid levels but these costs are not included in the economic model. The ERG notes this could slightly underestimate the cost of TOC treatment.

The number of rheumatologist visits was taken from TA375,[24] which assumed 10 visits during the first six months of treatment and monthly visits thereafter regardless of treatment.

Table 37: Administration and monitoring resource unit costs

Item	Unit cost (2015 £)	Source
Rheumatologist visit	£137.00	NHS Reference Costs 2014 to 2015:[74] WF01A
GP visit	£65.00	PSSRU 2015[75] (p. 177, 10.8b)
Nurse visit	£75.00	PSSRU 2015[75] (p. 172, 10.4)
Hospital day - Palliative	£371.00	PSSRU 2015[75] (p. 107, 7.1)
IV administration	£173.60	NICE TA247,[26] adjusted for inflation
Full blood count (FBC)	£3.01	NHS Reference Costs 2014 to 2015:[74] DAPS05
Urea and electrolytes (U&E)	£1.19	NHS Reference Costs 2014 to 2015:[74] DAPS04
Liver function test (LFT)	£3.01	NHS Reference Costs 2014 to 2015:[74] DAPS05
Creatinine (CRE)	£3.01	NHS Reference Costs 2014 to 2015:[74] DAPS05
Chest X-ray (CXR)	£30.23	NHS Reference Costs 2014 to 2015:[74] DAPF

Table 38: Monitoring schedules for different therapies

Treatment	Pre-treatment	On treatment	
		First 6 months	Subsequent 6 months
MTX*	FBC, U&E, LFT, CXR	11 x (FBC + U&E + LFT)	6.5 x (FBC + U&E + LFT)
Leflunomide	FBC, U&E, LFT, CRE	6.5 x (FBC + LFT)	3.25 x (FBC + LFT)
Ciclosporin	FBC, U&E, LFT, 2xCRE	6.5 x (FBC + LFT) 13 x (U&E + CRE)	2.16 x (FBC + LFT) 6.5 x (U&E + CRE)
Azathioprine	FBC, U&E, LFT	12 x (FBC + LFT), U&E + CRE	U&E, CRE
Sulfasalazine	FBC, U&E, LFT, CRE	4 x (FBC + LFT)	2.16 x (FBC + LFT)

*CZP, ABA, ADA, ETA, GOL, IFX, TOC, RTX, MTX, HCQ, NBT and gold injections assumed to have same schedule as MTX.
FBC= full blood count; U&E= urea and electrolytes; LFT= liver function test; CXR= chest x-ray; CRE= creatinine.

5.2.8.3 Health state costs

The company presented costs per HAQ band calculated by combining data on inpatient days and joint replacements from the Norfolk Arthritis Register (NOAR) database[76] with NHS Reference Costs,[77] which is the same approach taken by the assessment group in TA375.[24] Table 39 shows the direct costs associated with different bands of the HAQ score and the costs adjusted for inflation using the health component of the UK consumer price index.[78] Indirect costs are considered in one of the scenario analyses undertaken by the company and described in Section 5.2.9.3.

Table 39: Direct costs associated with different HAQ scores

HAQ score	Costs reported (2010)	Costs adjusted for inflation (2015)
<0.6	£167.41	£188.72
0.6 - 1.1	£102.54	£115.59
1.1 - 1.6	£364.68	£411.10
1.6 - 2.1	£523.68	£590.34
2.1 - 2.6	£1,246.26	£1,404.89
≥2.6	£2,687.97	£3,030.10

5.2.8.4 Other costs

The company's economic analysis did not include the costs associated with managing AEs and justified this exclusion on the basis that the safety profile of CZP is comparable to that of other bDMARDs.

The cost of palliative care was calculated from the resource use based on consultation with an expert rheumatologist. The expert rheumatologist estimated that patients receiving palliative care would require two rheumatologist visits at least every two months and treatment with prednisolone requiring admission to a day-care setting at least three times a year. Based on these assumptions the treatment cost of palliative care per 6-month period was estimated to be £978.

5.2.9 Cost effectiveness results

The ERG notes that during the clarification round the company identified and fixed two minor errors that affected the results of the economic analysis that, according to the company, had "limited impact on the estimated ICERs" (see Appendix 3 of the clarification response,[28]). The company provided the results for the base case analysis produced with the revised model along with the clarification responses, but not for the sensitivity analyses, stating that "sensitivity analyses will be similar to those presented in the original submission, given that there were only minor differences in base case results between the original and revised models". The ERG reproduces here the results of the base case analysis produced by the revised model and the results of the sensitivity analyses reported in the CS.[31]

5.2.9.1 Base case analysis

This section reproduces the results of the company's base case analysis. The base case analysis is applied to the three populations defined in Section 5.2.2. These populations are defined conditional on whether patients are contraindicated, or have had an AE, to MTX and/or RTX and are:

A. Adults eligible for RTX + MTX.
B. Adults for whom RTX is contraindicated or withdrawn due to an AE.
C. Adults for whom RTX therapy cannot be given because MTX is contraindicated or withdrawn due to an AE.

For each population, deterministic and probabilistic incremental analyses are presented along with cost effectiveness acceptability curves (CEACs). Probabilistic analyses are based on 5,000 Monte Carlo simulations and the mean costs and QALYs are presented. In addition to the probabilistic ICER, the probabilities of each competing option being cost-effective at thresholds of £20,000 and £30,000 per QALY gained are presented.

Within Population A, the probabilistic version of the company's model suggests that CZP+MTX before RTX+MTX is expected to generate an additional 0.29 QALYs at an additional cost of £9,647 compared with the same sequence without CZP+MTX; the corresponding ICER is expected to be £33,222 per QALY gained. Table 40 and Table 41 present the deterministic and probabilistic results, respectively. Figure 19 shows the CEAC of CZP + MTX before RTX + MTX.

In Population B, the probabilistic version of the company's model suggests that CZP + MTX was is expected to generate an additional 0.251 QALYs at an additional cost of £833 compared with ADA + MTX, resulting in an ICER of £3,317 per QALY gained. On the other hand, TOC (IV) + MTX is expected to generate an additional 0.201 QALYs at an additional cost of £26,658 compared with CZP + MTX resulting in a probabilistic ICER of £132,783 per QALY gained. However, this analysis excludes the CIC PAS associated with TOC and ABA. Table 42 and Table 43 present the deterministic and probabilistic results, respectively. Figure 20 shows the CEACs considering all of the bDMARDs simultaneously.

In Population C, the probabilistic version of the company's model suggests that CZP monotherapy is expected to generate an additional 0.274 QALYs at an additional cost of £1,352 compared with ADA monotherapy, resulting in an ICER of £4,943 per QALY gained. On the other hand, TOC (IV) monotherapy is expected to generate an additional 0.203 QALYs at an additional cost of £26,199 compared with CZP monotherapy, resulting in a probabilistic ICER of £129,177 per QALY gained. However, this analysis excludes the CIC PAS associated with TOC and ABA.

Table 44 and

Table 45 present the deterministic and probabilistic results, respectively. Figure 21 shows the CEACs considering all bDMARD monotherapies.

Table 40: Deterministic results of the base case analysis results for patients eligible for RTX + MTX (Population A)

Sequences	Total QALYs	Total costs	Inc. QALYs	Inc. costs	ICER (£/QALY)
RTX ‡	7.000	£138,520	-	-	-
CZP before RTX†	7.286	£148,361	0.286	£9,842	**£34,378**

‡ RTX: RTX+MTX, TOC+MTX, ABA+MTX, MTX + HCQ + SSZ, NBT, PC
† CZP before RTX: CZP+MTX, RTX+MTX, TOC+MTX, ABA+MTX, MTX + HCQ + SSZ, NBT, PC

Table 41: Probabilistic results of the base case analysis results for patients eligible for RTX + MTX (Population A)

Sequences	Total QALYs	Total costs	Inc. QALYs	Inc. costs	ICER (£/QALY)	Probability (%) of cost effectiveness at a threshold of	
						£20,000 /QALY	£30,000 /QALY
RTX †	7.031	£139,933	-	-	-	97.80	63.02
CZP before RTX†	7.321	£149,579	0.290	£9,647	**£33,222**	2.20	36.98

‡ RTX: RTX+MTX, TOC+MTX, ABA+MTX, MTX + HCQ + SSZ, NBT, PC
† CZP before RTX: CZP+MTX, RTX+MTX, TOC+MTX, ABA+MTX, MTX + HCQ + SSZ, NBT, PC

Figure 19: Cost effectiveness acceptability curve for patients eligible for RTX + MTX (Population A) (reproduced from Figure D of the company's clarification response)[28]

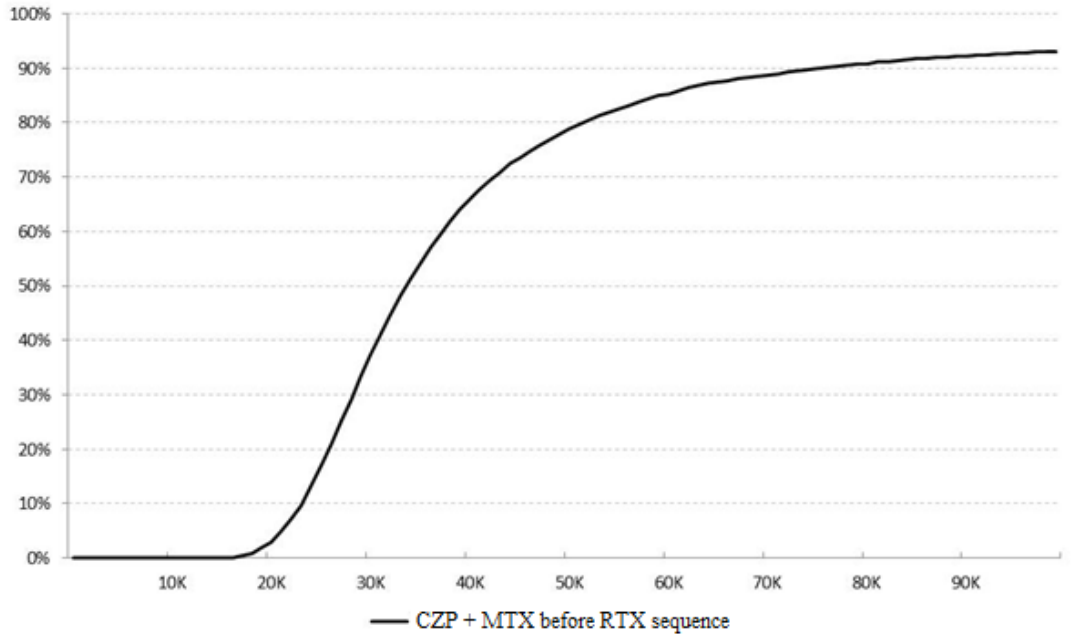


Table 42: Deterministic results of the base case analysis results for patients for whom RTX is contraindicated or withdrawn (Population B)

First therapy of the sequence†	Total QALYs	Total costs	Inc. QALYs	Inc. costs (£)	ICER (£/QALY)
IFX + MTX	6.048	£101,484	-	-	Dominated
ETA + MTX	6.048	£97,606	-	-	Dominated
ADA + MTX	6.048	£97,183	-	-	-
GOL + MTX	6.048	£97,183	-	-	-
ABA(IV) + MTX‡	6.095	£115,555	0.047	£18,373	Dominated
CZP + MTX	6.308	£98,100	0.260	£918	**£3,527**
TOC(IV) + MTX‡	6.507	£125,112	0.199	£27,011	**£135,953**

†Rest of the sequence: MTX + HCQ + SSZ, LEF, GLD, CIC, AZA, PC
‡CiC PAS not included

Table 43: Probabilisitic results of the base case analysis results for patients for whom RTX is contraindicated or withdrawn (Population B)

First therapy of the sequence†	Total QALYs	Total costs	Inc. QALYs	Inc. costs	ICER (£/QALY)	Probability(%) of cost effectiveness at a threshold of	
						£20,000/ QALY	£30,000/ QALY
IFX + MTX	6.038	£102,242	-	-	Dominated	0.00	0.00

ETA + MTX	6.070	£98,360	-	-	Dominated	0.0	0.7
GOL + MTX	6.071	£97,964	-	-	-	0.3	1.5
ADA + MTX	6.076	£98,015	-	-	Extendedly dominated	0.2	1.7
ABA (IV)+ MTX‡	6.119	£116,232	-	-	Dominated	0.00	0.00
CZP + MTX	6.327	£98,848	0.256	£884	**£3,461**	99.5	96.0
TOC (IV)+ MTX‡	6.528	£125,507	0.201	£26,659	**£132,783**	0.00	0.00

† Rest of the sequence: MTX + HCQ + SSZ, LEF, GLD, CIC, AZA, PC
‡ CiC PAS not included

Figure 20: Cost effectiveness acceptability curves for patients for whom RTX is contraindicated or withdrawn (Population B) (reproduced from Figure E of the company's clarification response)[28]

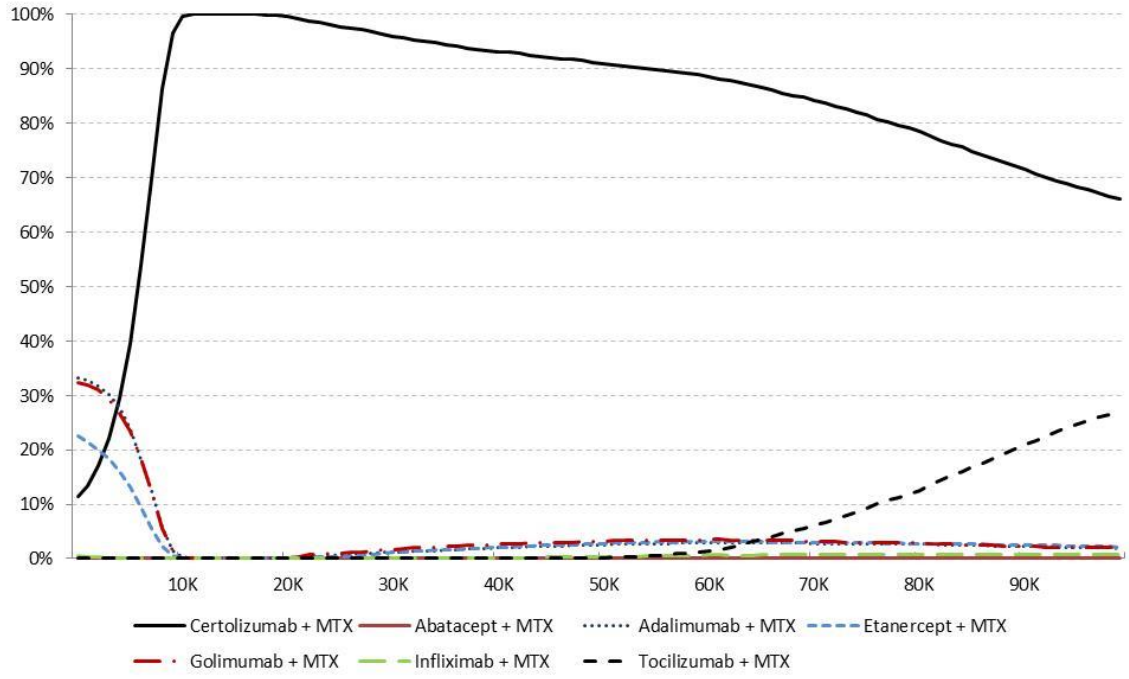


Table 44: Deterministic results of the base case analysis results for patients for whom MTX is contraindicated or withdrawn (Population C)

First therapy of the sequence†	Total QALYs	Total costs	Inc. QALYs	Inc. costs	ICER (£/QALY)
ADA	5.880	£95,632	-	-	-
ETA	5.880	£96,036	-	-	Dominated
CZP	6.141	£97,249	0.260	£1,617	**£6,213**
TOC (IV) ‡	6.346	£123,592	0.206	£27,960	**£127,955**

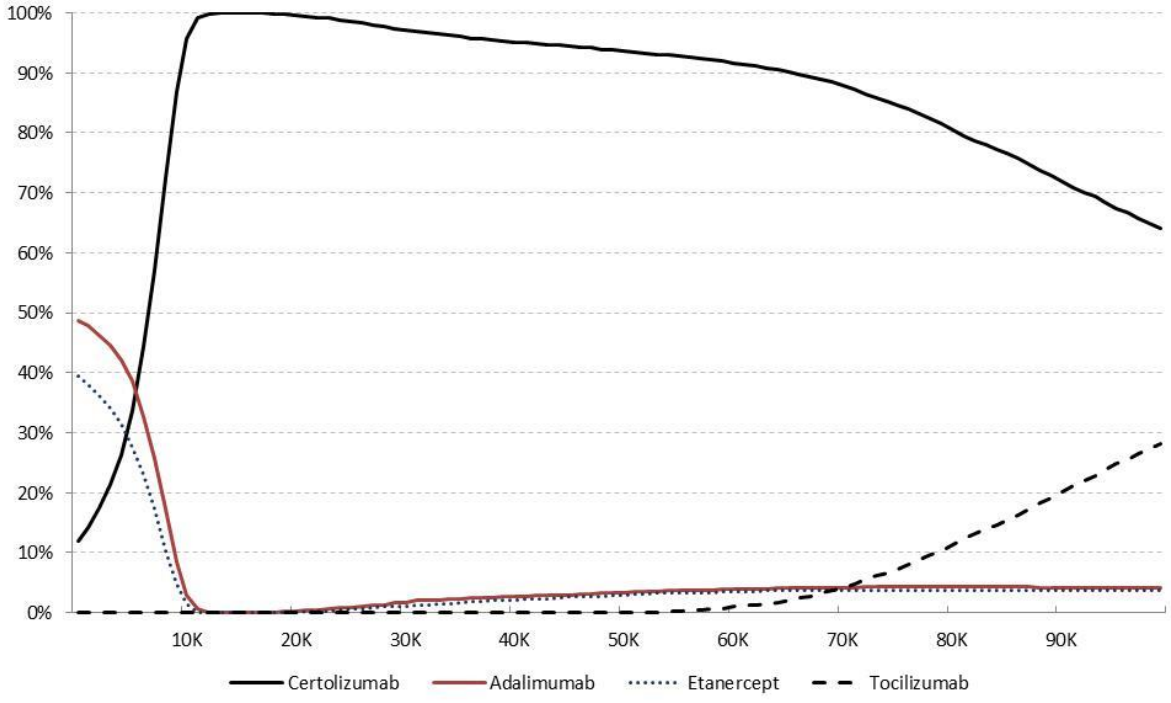
† Rest of the sequence: LEF, GLD, CIC, AZA, PC
‡ CiC PAS not included

Table 45: Probabilistic results of the base case analysis results for patients for whom MTX is contraindicated or withdrawn (Population C)

First therapy of the sequence†	Total QALYs	Total costs	Inc. QALYs	Inc. costs	ICER (£/QALY)	Probability (%) of cost effectiveness at a threshold of	
						£20,000 /QALY	£30,000/ QALY
ETA	5.899	£96,270	-	-	Dominated	0.04	0.92
ADA	5.902	£95,918	-	-	-	0.18	1.16
CZP	6.162	£97,254	0.260	£1,336	**£5,151**	99.78	97.48
TOC(IV) ‡	6.358	£123,433	0.196	£26,179	**£133,655**	0.00	0.00

† Rest of the sequence: LEF, GLD, CIC, AZA, PC
‡ CiC PAS not included

Figure 21: Cost effectiveness acceptability curves for patients for whom MTX is contraindicated or withdrawn (Population C) (reproduced from Figure F of the company's clarification response) [28]



5.2.9.2 One-way sensitivity analyses

The company performed a series of deterministic one-way sensitivity analyses to test the impact of parameters on the outcomes of the model. The parameters and the variation applied to them are summarised in Table 46.

Table 46: Parameters included in the one-way sensitivity analysis and the and ranges evaluated

Parameter	Variation
Discount rates for costs and effects	0-6%
Mean baseline HAQ	30 % variation
Mean baseline pain	30 % variation
Mean baseline EQ-5D	30 % variation
Trial-specific baseline effects in the NMA model[a]	95% CrI
Cut-off statistics (Z) in the NMA model (see Section 5.2.6.1)[b]	95% CrI
HAQ mortality hazard ratio	95% CrI
Coefficient of HAQ for the mapping to EQ-5D	30% variation
Effect of CZP treatment on probability of EULAR response	95% CrI
Effect of comparator treatment on probability of EULAR response	95% CrI

[a] Assumed by the ERG to mean the "No response" rate from the NMA for the reference treatment in the REALISTIC[34] study
[b] Assumed by the ERG to mean the common value across studies included in the NMA that splits responders between moderate and good responders for the reference treatment

The results of the deterministic one-way sensitivity analysis are presented as tornado diagrams in Figure 22, Figure 23, and Figure 24 for Populations A, B and C, respectively. The ERG produced these diagrams using the company's revised model following the clarification process. The horizontal axis shows the percentage change in net monetary benefit assuming a cost per QALY threshold of £30,000. The ERG notes that the interpretation of the tornado diagrams is not straight forward because they show the percentage change from the base case NMB without showing the baseline NMBs.

The CS stated that the parameters that exhibited the most substantial impact on results were the treatment effects (of CZP and comparators) on the probability of EULAR response, the discounting rates for costs and effects and the coefficient of HAQ for the mapping to EQ-5D. Variation in mean baseline HAQ, pain and EQ-5D, the mortality hazard ratio associated with a unit increase in HAQ, trial-specific baseline effects (which the ERG assumes to mean the "No response" rate from the NMA for the reference treatment in the REALISTIC[34] study) and the cut-off statistics (assumed by the ERG to mean the common value across studies included in the NMA that splits responders between moderate and good responders for the reference treatment) were found to have a smaller impact on results.

Figure 22: Tornado diagram showing the percentage change in net monetary benefit at a willingness to pay threshold of £30,000 per QALY gained, based on variation in individual model parameters – Population A (produced using the revised model)

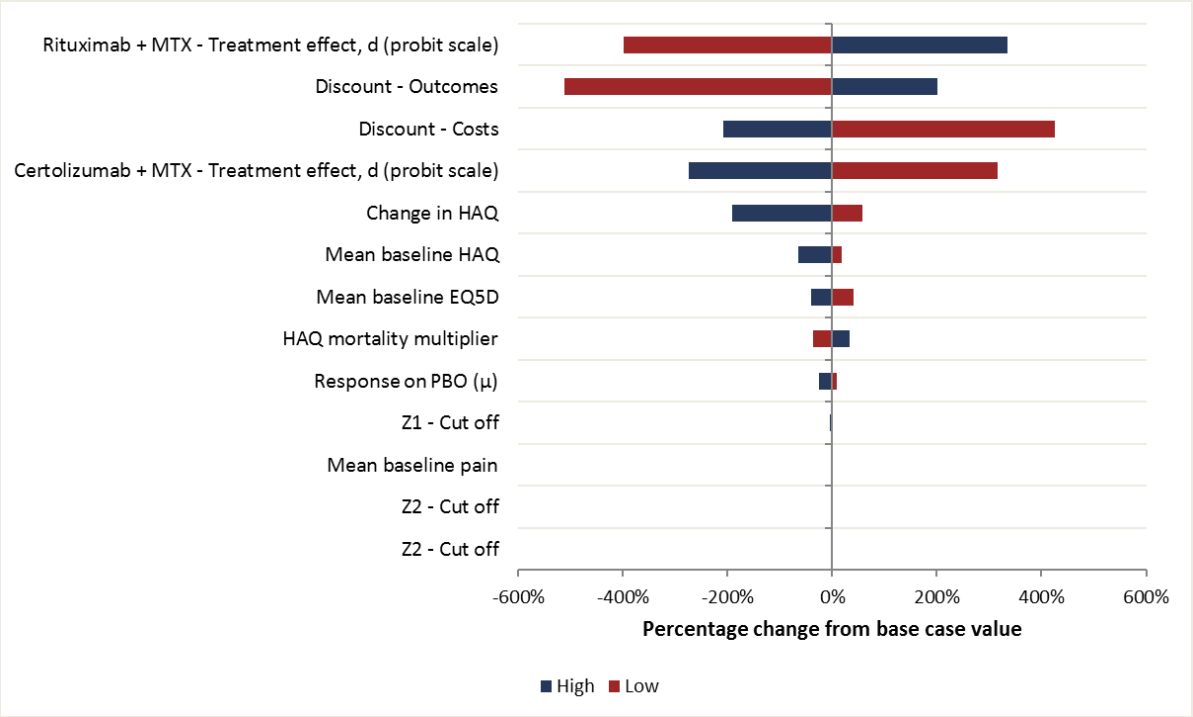


Figure 23: Tornado diagram showing the percentage change in net monetary benefit at a willingness to pay threshold of £30,000 per QALY gained, based on variation in individual model parameters – Population B (CZP+MTX versus TOC+MTX) (produced using the revised model)

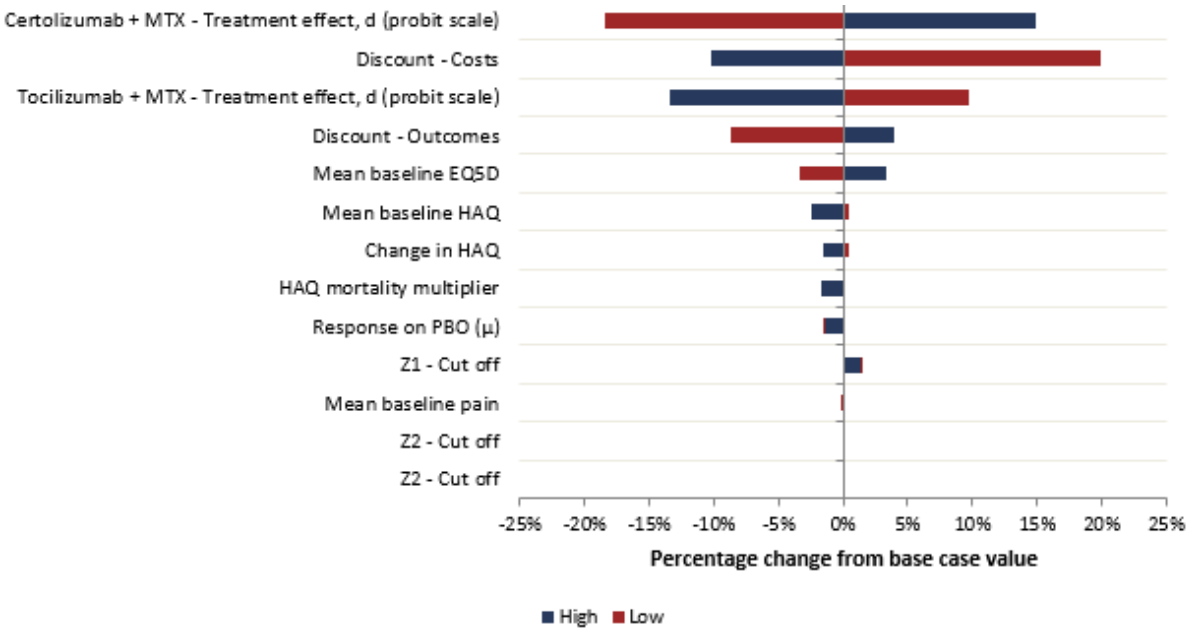
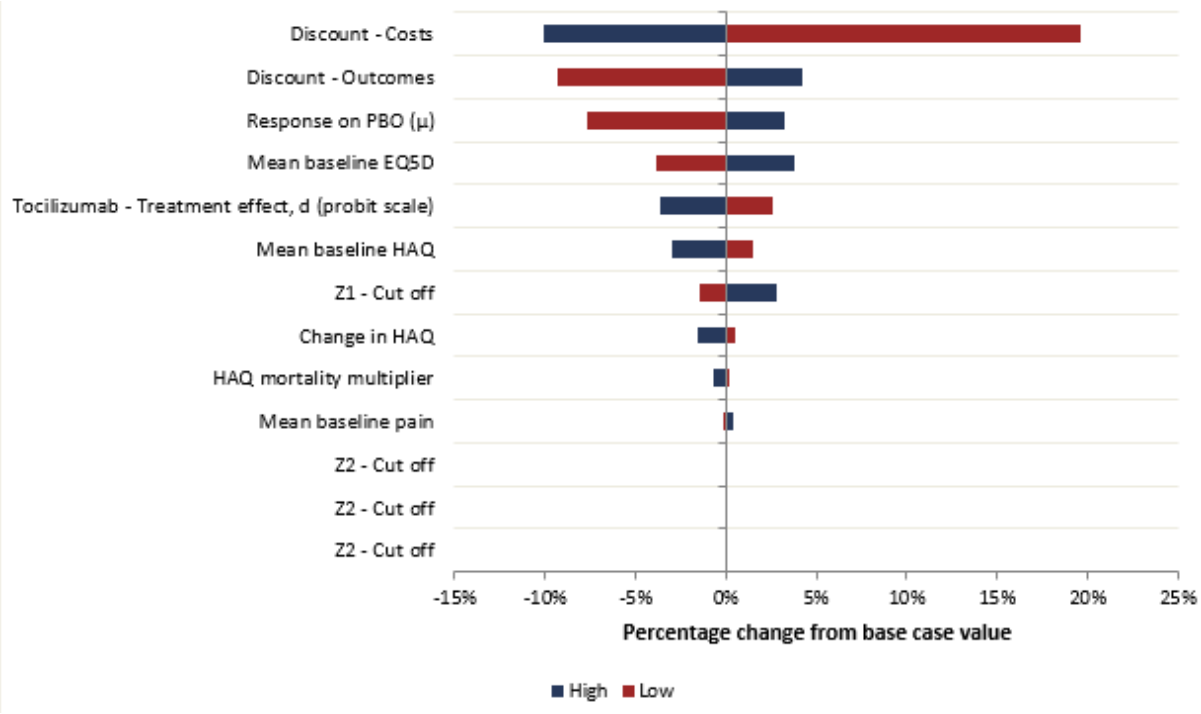


Figure 24: Tornado diagram showing the percentage change in net monetary benefit at a willingness to pay threshold of £30,000 per QALY gained, based on variation in individual model parameters – Population C (CZP versus TOC) (produced using the revised model)



5.2.9.3 Scenario analyses

The company undertook a series of scenario analyses to explore the impact of the base case assumptions on the ICER. The parameters and their respective alternative assumptions explored in the scenario analyses are listed in Table 47.

For the base case, the efficacy of CZP was calculated using the NMA as described in Section 4.3. However, the company undertook different scenario analyses using alternative assumptions regarding relative effectiveness. In one scenario analysis, the efficacy of CZP + MTX (used in Populations A and B) was assumed to be equal to that of GOL + MTX, and therefore also equal to the rest of the TNFis. For Population C, the efficacy of ADA and ETA monotherapies were assumed to be equal to that of CZP monotherapy. In another scenario analysis, the results of the NMA including the J-RAPID study were used to inform the probabilities of EULAR response. The company excluded the J-RAPID study from the NMA for their base case analysis "because of small sample size and the associated risk of a biased effect size estimate." (p.183 of the CS)[31]

The retreatment interval of RTX is subject to uncertainty. The Appraisal Committee of TA195 concluded that the 8.7-month retreatment interval assumed by the Assessment Group was likely to

overestimate the time between consecutive courses of rituximab. However, based on clinical specialists' advice, the Committee considered it was unlikely that the mean retreatment interval would be as low as 6 months. The company assumed a 6-month retreatment interval for the base case analysis but explored the impact on the ICER of a 9-month retreatment interval in a scenario analysis as assumed in TA375.[24]

The company's model uses a mapping from HAQ score to EQ-5D. In the base case, a linear mapping attributed to Brennan *et al.*[70] is used which maps changes in HAQ to changes in EQ-5D by multiplying them by -0.2102. The company undertook a scenario analysis using the algorithm proposed by Hernández Alava *et al.* [65] but due to issues in its implementation (explained in Section 5.3) the ERG decided not to replicate these results in this report.

In order to estimate the change in utility from baseline on initial response to the first therapy considered in the analysis (i.e., second bDMARD), the company fitted a linear regression model to the data from the PREDICT study. A scenario analysis explored the impact on the ICER of estimating this utility using the same mapping from HAQ to EQ-5D as in the rest of the model.

The company assumed within the base case analysis that the patients had reached the utility gain associated with their 6-month EULAR response by week 6. A scenario analysis undertaken by the company explores the impact of assuming a percentage of the utility gain is achieved at week 6 with the remaining utility gained at a constant rate between week 6 and month 6.

The base case analysis assumes that the time to treatment discontinuation for the first treatment is independent of bDMARD (i.e. that the times are equal for all biologics). The company explored the impact of assuming different values for the scale parameter of the Weibull distribution for TNFi (0.3003) and non-TNFi treatments (0.2208), allowing these classes to have differential time to discontinuation. The source of this value was not provided in the CS.[31]

The CS also includes a scenario analysis that the company claims uses a societal perspective. The only difference with the base case analysis is the inclusion of indirect costs associated with patients' HAQ scores. Indirect costs per HAQ band were taken from a paper reporting costs from the Early RA Study (ERAS).[79] A breakdown of the indirect costs per HAQ score band is provided in Table 48.

Other scenario analyses undertaken by the company assumed: patients on cDMARDs and palliative care would experience no HAQ progression; no vial wastage; and removing the maximum value of 2.76 for HAQ.

Table 47: Parameters and alternative assumptions explored in the scenario analyses

Parameter		Base case assumption	Alternative assumption(s)
Efficacy of CZP		Using results from the NMA (Efficacy of CZP taken from the REALISTIC study)	For Populations A and B, assume CZP has same efficacy as the rest of TNFis. For Population C, assume ADA and ETA have same efficacy as CZP
			Including J-RAPID in the NMA
Retreatment interval of RTX		6 months	9 months
Mapping from HAQ to EQ-5D		Using coefficient (-0.2102) attributed to Brennan *et al.*[70]	Using pain and HAQ, based on Hernández Alava *et al.* [65]
Estimates of utility improvements on initial response to first-line treatment		Linear regression model fitted to data from the PREDICT study	Change from baseline utility mapped from change in HAQ score
% of patients enjoying full utility gains after six weeks of first treatment		100%	25%
Time to treatment discontinuation of first therapy (scale parameter of Weibull distribution)	Non-TNFi	0.4416	0.2208
	TNFi	0.4416	0.3003
Perspective		NHS/PSS	Societal
HAQ progression on cDMARDs		0.045 increase per year	0.000 increase per year (i.e. no change in HAQ)
HAQ progression on palliative care		0.06 increase per year	0.000 increase per year (i.e. no change in HAQ)
Vial wastage		Yes	No
Maximum HAQ score		2.76	3.0

Table 48: Costs associated to each HAQ category including indirect costs

HAQ category	Total costs reported (2001 USD)	Costs adjusted for currency and inflation (2016 £)
<0.6	$221	£189.62
0.6 - 1.1	$3767	£3,232.09
1.1 - 1.6	$5,185	£4,448.73
1.6 - 2.1	$7,910	£6,786.78
2.1 - 2.6	$12,045	£10,334.61
≥2.6	$12,548	£10,766.18

Exchange rate applied £1.00 = $1.58 (Q4 2015 average), source for inflation index not provided in CS

The results of the scenario analyses are reproduced in Table 49. The ERG notes that the results of the scenario analyses for Population C contain some errors; these are described in Section 5.3.

Table 49: Results of scenario analyses undertaken by the company, incremental analyses showing ICERs (£/QALY) (deterministic)

Parameter	Base case	Scenario analysis	Population A	Population B (Incremental analysis)							Population C (Incremental analysis)			
			CZP + MTX vs RTX + MTX	CZP + MTX	ABA + MTX†	ADA + MTX	ETA + MTX	GOL + MTX	IFX + MTX	TOC + MTX†	CZP	ADA	ETA	TOC†
Base case analysis			**£34,516**	**£3k**	**D**	**-**	**D**	**-**	**D**	**£129k**	**£5k**	**-**	**D**	**£123k**
Source of utility for first treatment‡ response	Linear regression (PREDICT)	HAQ score from REALISTIC mapped to EQ-5D	£33,199	£6k	D	-	D	-	D	£204k	£8k	-	D	£189k
% patients enjoying utility gain at 6 weeks	100%	25%	£34,430	£3k	D	-	D	-	D	£132k	£5k	-	D	£126k
Efficacy of CZP	Based on NMA	For Populations A and B, assume CZP efficacy equal to other TNFis. For Population C, assume ADA and ETA efficacy equal to CZP's	£169,690	-	ED	D	D	D	D	£62k	-	D	D	£793k
		Including J-RAPID in the NMA	£29,613	£4k	D	-	D	-	D	£182k	£7k	-	D	D
Duration of non-TNF therapy (scale parameter, Weibull)	0.4416	0.2208	D	£4k	ED	-	D	-	D	£43k	£5k	-	D	£44k
Duration of TNF therapy (scale parameter, Weibull)	0.4416	0.3003	£19,673	£7k	D	-	D	-	D	£2M	£7k	D	D	D
Vial wastage	Yes	No	£34,110	£4k	D	-	D	-	D	£98k	£5k	-	D	£94k
RTX retreatment interval	6 months	9 months	£49,618	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Perspective	NHS/PSS	Societal	£4,729	-	D	D	D	D	D	£118k	£5k	-	D	£135k
HAQ progression on cDMARDs	0.045 per annum	0 per annum	£53,578	£5k	D	-	D	-	D	£140k	£5k	-	D	£133k
HAQ progression on palliative care	0.06 per annum	0	£57,156	£7k	D	-	D	-	D	£155k	£10k	-	D	£155k
Maximum mean HAQ	2.76	3.0	£34,183	£4k	D	-	D	-	D	£130k	£5k	-	D	£123k

‡=first treatment considered in model (i.e., second bDMARD);†=CiC PAS not included; - = baseline; D = dominated; ED = extendedly dominated; NA=not applicable;

Within further scenario analyses, the company included additional comparators: the use of subcutaneous (SC) formulations of TOC and ABA, and IFX biosimilars (Inflectra and Remsima). The cost of these drugs, along with their dosing regimens are included in Table 35. The SC formulations have the advantage of excluding the IV administration costs. The ERG chose to show the results of a full incremental analysis calculated using the company's model including all formulations. The results for Populations B and C are shown in Table 50 and Table 51, respectively. The ERG notes that the model contains an error (described in Section 5.3) that leads to the overestimation of the cost for TOC (SC) monotherapy in Population C (Table 51).

Table 50: Incremental analysis for Population B including SC formulations of TOC and ABA and IFX biosimilars (deterministic)

First therapy of the sequence†	Total QALYs	Total costs	Inc. QALYs	Inc. costs	ICER (£/QALY)
GOL + MTX	6.016	£97,593	-	-	-
ADA + MTX	6.016	£97,593	£0	£0	-
ETA + MTX	6.016	£98,017	0.000	£423	Dominated
IFX (bio) + MTX	6.016	£99,086			Dominated
IFX + MTX	6.016	£101,894			Dominated
ABA (SC) + MTX	6.065	£115,609			Dominated
ABA (IV) + MTX‡	6.065	£118,410			Dominated
CZP + MTX	6.286	£98,575	0.270	£981	**£3,641**
TOC (SC) + MTX	6.491	£112,716	0.205	£14,141	**£68,953**
TOC (IV) + MTX‡	6.491	£125,096	0	£12,380	Dominated

†Rest of the sequence: MTX + HCQ + SSZ, LEF, GLD, CIC, AZA, PC
‡CiC PAS not included
bio = biosimilar

Table 51: Incremental analysis for Population C including TOC (SC) (deterministic)

First therapy of the sequence†	Total QALYs	Total costs	Inc. QALYs	Inc. costs	ICER (£/QALY)
ADA	5.845	£95,943	-	-	-
ETA	5.845	£96,347			Dominated
CZP	6.115	£97,292	0.271	£1,349	**£4,985**
TOC (SC)	6.328	£123,695	0.213	£26,403	**£123,915**
TOC (IV)‡	6.328	£145,418			Dominated

† Rest of the sequence: LEF, GLD, CIC, AZA, PC
‡ CiC PAS not included

In addition, the company analysed the impact of a shorter time horizon (5-years and 10-years) and different discount rates (the four combinations of 1.5% and 6.0% for costs and effects) although the ERG does not believe that these analyses should be preferred to those in the base case. For the sake of brevity, the results of these analyses are not reproduced in this report. The full results of the scenario analyses undertaken by the company are included in Table 101 of the CS.

In response to the ERG's clarification questions (see clarification response,[28] question B21), the company provided a scenario analysis where they used the change in HAQ conditioned on EULAR response status used in the AG's analysis for TA375:[24] 0 for no responders, -0.317 for moderate responders and -0.672 for good responders. The results of this scenario analysis are presented in Table 52, Table 53 and

Table 54 for populations A, B, and C respectively.

Table 52: Scenario analysis results using HAQ score changes from TA375 for Population A (deterministic)

Treatment	Total QALYs	Total costs	Inc. QALYs	Inc. costs	ICER (£/QALY)
RTX + MTX	6.998	£138,186			
CZP+ MTX	7.284	£148,105	0.286	£9,919	**£34,635**

‡ RTX: RTX+MTX, TOC+MTX, ABA+MTX, MTX + HCQ + SSZ, NBT, PC
† CZP before RTX: CZP+MTX, RTX+MTX, TOC+MTX, ABA+MTX, MTX + HCQ + SSZ, NBT, PC

Table 53: Scenario analysis results using HAQ score changes from TA375 for Population B (deterministic)

Treatment	Total QALYs	Total costs	Inc. QALYs	Inc. costs	ICER (£/QALY)
GOL + MTX	6.045	£95,976			
ADA + MTX	6.045	£95,976	0.000	£0	-
ETA + MTX	6.045	£96,400	-	-	Dominated
IFX + MTX	6.045	£100,277	-	-	Dominated
ABA + MTX	6.092	£114,399	-	-	Extendedly dominated
CZP + MTX	6.305	£97,181	0.260	£1,205	**£4,637**
TOC + MTX	6.503	£124,371	0.198	£27,190	**£137,138**

†Rest of the sequence: MTX + HCQ + SSZ, LEF, GLD, CIC, AZA, PC
‡CiC PAS not included

Table 54: Scenario analysis results using HAQ score changes from TA375 for Population C (deterministic)

Treatment	Total QALYs	Total costs	Inc. QALYs	Inc. costs	ICER (£/QALY)
ADA	5.874	£94,542			
ETA	5.874	£94,946	-	-	Dominated
CZP	6.133	£96,390	0.259	£1,848	**£7,127**
TOC	6.338	£122,859	0.205	£26,469	**£129,102**

† Rest of the sequence: LEF, GLD, CIC, AZA, PC
‡ CiC PAS not included

In response to another clarification question, (see clarification response,[28] question B21), the company provided the results of a scenario analysis where mortality was not affected by change in HAQ score, but only by the baseline HAQ score, as assumed by the AG in TA375.[24] However, the company used an estimated multiplier for all-cause mortality of 1.87, instead of 1.43^1.55=1.74 and therefore, the results reported are of limited validity.

5.3 Critical appraisal of the company's economic evaluation

5.3.1 Methods for reviewing the company's economic evaluation and health economic model

The ERG adopted a number of approaches to explore, interrogate and critically appraise the company's submitted economic evaluation and the underlying health economic model upon which this was based. These approaches included:

- Consideration of key items contained within published economic evaluation and health economic modelling checklists to critically appraise the company's model and analysis.[80, 81]
- Scrutiny of the company's model by health economic modellers including:
  - White-box validation: checking of inputs, code and formulae
  - Black-box testing: changing inputs to check whether the output matches expectations
  - Face-validity testing: checking model results match expectations
  - Comparison of deterministic and probabilistic ICERs.
- Replication of the base case results, PSA and scenario analysis presented within the CS.[31]
- Where possible, checking parameter values used in the company's model against the original data sources.
- Examination of concordance between the description of the model reported within the CS[31] and the company's executable model.
- The use of expert clinical input to judge the clinical robustness of the company's economic evaluation and of the assumptions underpinning the model.

5.3.2 Summary of main limitations identified within the critical appraisal

The main potential limitations identified within the ERG's critical appraisal of the company's economic analysis are described under the following headings:

1. Deviations from the NICE Reference Case
2. Appropriateness of sequences compared for Population A
3. Appropriateness of including ABA + MTX therapy after TOC + MTX
4. Appropriateness of the methods used for the NMA
5. Exclusion of the J-RAPID trial from the NMA
6. Modelling of HAQ progression on cDMARDs and palliative care
7. Modelling of HAQ to EQ-5D mapping
8. Retreatment interval of RTX
9. Appropriateness of assuming treatment duration of TNFis is equal to that of other bDMARDs
10. Appropriateness of assuming changes in HAQ score affect mortality
11. Failure to age-adjust utilities
12. Modelling of HAQ improvement in responders for subsequent therapies
13. Modelling of treatment discontinuation for subsequent therapies
14. Inaccuracy in TOC (IV) dosing
15. Approximation of the weight distribution of the population using weight bands
16. Inconsistency in benefits of treatment response during the first cycle
17. Exclusion of AEs
18. Inaccuracies in the number of doses per cycle
19. Appropriateness of using EQ-5D data from the PREDICT study
20. Perceived model errors and other issues surrounding model implementation

(1) Deviations from NICE Reference case

Table 55 summarises the extent to which the company's model adheres to the NICE Reference Case.[62]

Table 55: Adherence of the company's economic analysis to the NICE Reference Case

Element	Reference case	ERG comments
Defining the decision problem	The scope developed by NICE	The scope of the company's model is generally in line with the final NICE scope.[27] The population considered directly relates to the TNFi-experienced population of the REALISTIC study.[34] Clinical advice received by the ERG suggests that this is likely to be reflective of the UK population of TNFi-experienced adults with moderate to severe active RA population who may be eligible for CZP.
Comparator(s)	As listed in the scope developed by NICE	The final NICE scope[82] defines RTX in combination with MTX as a comparator for adults treated with at least 1 TNFi (unless either RTX or MTX are contraindicated or have been withdrawn). However, the company has compared a sequence with CZP+MTX followed by RTX+MTX with a sequence with RTX alone. The ERG does not think this is an appropriate comparison. The scope also defined best supportive care (BSC) as a comparator for people with moderate to severe, active disease despite treatment with bDMARDs. The company claimed that this patient group did not reflect current NICE recommendations for TNFis and that limited evidence supports the evaluation of CZP compared with BSC. As such, the CS[31] does not include any assessment of CZP compared with best supportive care. The base case analysis does not include biosimilars and the use of SC formulations of TOC and ABA is not included.
Perspective on outcomes	All direct health effects, whether for patients or, when relevant, carers	Health gains for patients are modelled in terms of QALYs gained.
Perspective on costs	NHS and PSS	The CS[31] states that an NHS and PSS perspective was adopted, although no relevant PSS costs are included in the company's model. All costs are assumed to be incurred in the secondary care setting. A scenario analysis explores what the CS calls "the societal perspective" consisting of the inclusion of indirect costs associated with defined HAQ bands
Type of economic evaluation	Cost-utility analysis with fully incremental analysis	The company's economic evaluation takes the form of a cost-utility analysis. The results of the analysis are presented in terms of the incremental cost per QALY gained for CZP versus its comparators.
Time horizon	Long enough to reflect all important differences in costs or outcomes between the technologies being compared	The model adopts a lifetime horizon. Scenario analyses are also presented for shorter time horizons (5 years and 10 years).

Element	Reference case	ERG comments
Synthesis of evidence on health effects	Based on systematic review	The probabilities of EULAR response for the intervention and the comparators are largely based on an NMA performed using data identified through a systematic review. However, the ERG has concerns with the NMA (see Section 4.3.) The extrapolation of the efficacy of CZP + MTX compared to PBO + MTX from 3 to 6 months was performed through use of a mapping matrix data generated from data from the RAPID 1 and 2 trials.[63, 64] HAQ improvement for responders to the first treatment is based on data collected in the REALISTIC study.[34] The probabilities of EULAR response and the HAQ improvements conditional on the EULAR response were based on the RADIATE study.[40]
Measuring and valuing health effects	Health effects should be expressed in QALYs. The EQ-5D is the preferred measure of HRQoL in adults.	Health utilities were based on EQ-5D estimates from the PREDICT study[36] and HAQ scores collected in the REALISTIC study[34] which were mapped to EQ-5D scores.
Source of data for measurement of health-related quality of life	Reported directly by patients and/or carers	Health utilities were based on EQ-5D estimates reported by patients taking part in the PREDICT study[36] and HAQ scores elicited from patients in the REALISTIC study[34] which were mapped to EQ-5D scores.
Source of preference data for valuation of changes in HRQoL	Representative sample of the UK population	
Equity considerations	An additional QALY has the same weight regardless of the other characteristics of the individuals receiving the health benefit	No additional equity weighting is applied to the estimated QALY gains.
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	Resource use estimates associated to HAQ categories were based on data from the NOAR database[76] and resource use for palliative care was estimated by one expert rheumatologist. Cost estimates were based on the BNF,[72] NHS Reference Costs[74] and the PSSRU.[75]
Discount rate	The same annual rate for both costs and health effects (currently 3.5%)	All costs and QALYs are discounted at a rate of 3.5%

(2) Appropriateness of sequences compared for Population A

For patients who were eligible for RTX and MTX (Population A), the scope defined RTX + MTX as a comparator to CZP + MTX. The company's economic analysis for this population compared the two sequences described in Table 27, which are essentially the same sequence except for the fact that the CZP + MTX treatment has been added to the beginning of the sequence. This results in sequences of

different lengths being compared. During the clarification process (see clarification response,[28] question B2), the ERG asked the company why, if elongated sequences were being considered, CZP + MTX had not also been considered after RTX + MTX. The company replied that the reasons to do this were twofold: that "the decision as to whether CZP can be given as an alternative to therapies in patients withdrawn from RTX or MTX, is addressed separately in Populations B and C of the economic analysis; and that "given limitations in the evidence surrounding the efficacy of subsequent therapies, (…) the submitted model is not designed to simultaneously assess the optimal positioning of CZP in the RA treatment pathway" (see clarification response,[28] question B2). The ERG notes that Populations B and C do not address this decision because: having had a previous RTX treatment is not equivalent to RTX being contraindicated or withdrawn due to an AE (Population B), or to MTX being contraindicated or withdrawn (Population C), and these analyses do not allow a comparison of CZP + MTX followed by RTX + MTX versus RTX + MTX followed by CZP + MTX. The ERG also notes that the company's second argument, namely the lack of evidence surrounding the efficacy of subsequent therapies, could be seen as an argument against the comparison of sequences of different length: the efficacy of subsequent treatments might be underestimated or overestimated, which would bias the results in favour of shorter or longer sequences respectively. The ERG believes that the single analyses presented by the company is insufficient and that further sequences should be analysed and compared incrementally.

(3) Appropriateness of including ABA + MTX therapy after TOC + MTX

The ERG notes that ABA + MTX was included as the third and fourth line of therapy in the comparator and intervention sequences, respectively. This does not appear to be consistent with the interpretation of NICE guidance used in TA375,[24] hence the ERG sought clarification from the company. In response, the company highlighted NICE recommendations that stated that ABA + MTX was recommended in patients "who have had an inadequate response to, or have an intolerance of, other DMARDs, including at least one TNFi, and who cannot receive rituximab therapy because they have a contraindication to rituximab, or when rituximab is withdrawn because of an adverse event" (see clarification response,[28] question B3). The company claimed that the fact that RTX + MTX was before ABA + MTX in both sequences meant that by the time patients reach ABA + MTX treatment, they could not "receive rituximab therapy because they have a contraindication to rituximab, or when rituximab is withdrawn because of an adverse event." The ERG notes that according to the company's model, patients discontinue RTX treatment due to lack of response or loss of efficacy and not for contraindication or withdrawal due to an AE and therefore the quoted NICE recommendation is not applicable in this case. The ERG notes that if the NICE recommendation for ABA would be applicable in this case, other bDMARDs such as ADA, IFX and ETA could have been included after ABA in the sequence too. In addition, these bDMARDs could have also been included after the first therapy in the sequences compared in Population B. The company also refers to "external clinical

advice that in TNFi-IRs, abatacept would be provided after failure on rituximab and tocilizumab." However, clinicians consulted by the ERG suggested that after RTX + MTX, either TOC + MTX or ABA + MTX were used, but not both. The ERG considers that using TOC + MTX after RTX + MTX is more in line with NICE recommendations and TA375.[24]

(4) Appropriateness of the methods used for the NMA

The ERG considers the description of the NMA to be statistically imprecise and not necessarily what should have been done. According to the CS, the EULAR response probabilities were estimated using summary statistics on the trial-specific baseline effects, effect size estimates and cut-off statistics extracted from the NMA. Instead of "trial-specific baseline effects" the ERG assumes that the company meant to say using the REALISTIC[34] EULAR response rates. However, the description of the NMA in the CS[31] also suggests that it is only the EULAR no response rate from REALISTIC[34] that is being used. From this description, the proportion of responders is separated between moderate and good responses depending on the evidence for this split from the other studies; the ERG suggests that the EULAR no response, moderate response and good response rates should have all come from the REALISTIC[34] study to reflect the assumed response rates in the target patient population. It is not clear what role the other studies have in estimating the moderate and good response rates.

The ERG believes that posterior means and standard deviations were extracted from the NMA and that these were used to generate the required probabilities by assuming univariate normal distributions; using univariate normal distribution to represent uncertainty about parameters is an unnecessary approximation that fails to preserve the underlying joint distribution between parameters. During the clarification process (see clarification response,[28] question A21), the company was asked to clarify why approximations of results using marginal univariate normal distributions or multivariate normal distributions were preferred to using draws from the joint posterior distribution (i.e. CODA) which would have maintained correlation when characterising uncertainty on the inputs to the economic model. The ERG considers that the rationale provided by the company for their approach to be wrong and disagrees with the claim that ignoring correlation is "unlikely to significantly affect the expected probabilities of response in the model, and hence not materially impact on the expected ICER (either from the deterministic analysis or from the expectation of the probabilistic analysis) from the analysis."

The company's use of a fixed effect model rather than a plausible random effects model means that the company is answering the question whether the treatments had an effect in the studies included in the NMA and/or ignoring any potential heterogeneity in treatment effects between studies. Heterogeneity is expected and the ERG would have preferred to see a random effects model incorporating weakly informative prior information for the between-study standard deviation. In the

presence of heterogeneity the treatment effect from a random effects model does not represent the treatment effect in any specific patient population and it is recommended to make inferences based on the predictive distribution of the treatment effect in a future study.[56] Similarly, a purpose of the evidence synthesis is to characterise the uncertainty on the inputs to the economic model and it is recommended that the joint posterior distribution is based on the predictive distribution of the effect in a new study.[83] The estimates provided by the company's model will underestimate uncertainty.

(5) Exclusion of J-RAPID from the NMA

The company excluded the J-RAPID study from the NMA for their base case analysis "because of small sample size and the associated risk of a biased effect size estimate."[31] The ERG comments that small studies are not inherently more biased than larger studies but the effects are more variable. The ERG believes that the J-RAPID study should only be excluded if there is evidence to suggest that the study was of poor quality. Details relating to the impact of including of J-RAPID on the comparative efficacy of treatments are provided on page 192 of the CS.[31]

(6) Modelling of HAQ progression for cDMARDs and palliative care

The company assumes an annual increase of the HAQ score for patients on cDMARDs of 0.045 and of 0.06 for palliative care. Whilst this assumption was used in previous NICE appraisals,[25] recent evidence from Norton *et al.*[68] which was used in TA375,[24] shows that the progression is better approximated using non-linear models. Norton *et al.*[68] estimated HAQ progression in patients not receiving bDMARDs using data from patients recruited to the ERAS inception cohort study. This is a large (n=1,460), UK-based cohort which has long-term follow-up (up to 10 years). A growth mixture model approach was taken for the analysis of the data. These findings have since been corroborated in the NOAR data set with follow-up to 15 years and the Early Rheumatoid Arthritis Network data set. The use of constant rates instead of the non-linear progression estimated by Norton *et al.*[68] is likely to overestimate the benefits of bDMARDs. However, it is unclear how this would impact the cost-effectiveness of CZP compared with other bDMARDs.

(7) Modelling of HAQ to EQ-5D mapping

In the company's base case, changes in HAQ scores are mapped to EQ-5D utilities using an algorithm that simply assigns a change of -0.2102 in EQ-5D per unit change in HAQ, attributed to Brennan *et al.*[70] The company clarified that these regressions from Brennan *et al.*[70] were based on results from a study by Bansback *et al.*[84] (see clarification response,[28] question B10) However, Brennan *et al.*[70] claim that they use a "validated mapping which imputes the EQ-5D from all 42 components of the HAQ disability questionnaire data", hence it is unclear to the ERG how this mapping could be reduced to a single coefficient.

The ERG notes that recent approaches have been shown to achieve a better mapping, such as the approach published by Hernández Alava *et al*;[65] this algorithm is used one of the company's scenario analyses. The approach proposed by Hernández Alava *et al.*[65] is a four-class mixture model, which is the combination of four different distributions assigned to four latent classes. The combination is performed though a weighted average based on the probabilities of class membership. Explanatory variables predict both the relationship with EQ-5D and the probability of class membership. However, the ERG considers that the company's implementation of this approach in the model is inappropriate. Firstly, it only includes a subset of the variables (HAQ and pain) and excludes Age and quadratic variables (e.g. HAQ^2) which will lead to inaccuracy in the estimated value. Secondly, the distributions are not combined and instead the model offers the choice to perform the mapping with the distribution belonging to a specific class. In response to a request for clarification on the implementation of this mapping algorithm, the company acknowledged that "at the time of development, there was no clear approach to incorporating quadratic or squared terms (i.e. HAQ^2) of the Hernandez-Alava *et al.* mapping algorithm in a cohort-based model that calculates outcomes on expected values" and correctly stated that $E[HAQ^2]$ does not equal $(E[HAQ])^2$ (see clarification response,[28] question B11). The ERG notes that this is true not only for the mapping from HAQ to EQ-5D but also for any other non-linear function dependent on HAQ, such as HAQ associated costs and mortality. The company's response seems to be at odds with the assertion included in the CS stating that the company "are not aware that individual patient simulation has a particular benefit over the chosen approach." However substantial the divergence between $E[HAQ^2]$ and $(E[HAQ])^2$ is, the ERG believes it is likely to be smaller than the error introduced by not taking the variable into account at all. The reasoning provided not to implement the class membership probabilities was similar, claiming that the "probability of class membership conditional on the expected baseline characteristics of the cohort, (…) will differ to the expected probability of class membership conditional on patient-level baseline characteristics" (see clarification response,[28] question B11). The company pointed out the fact that applying the distributions independently provides a range of plausible ICERs. The ERG agrees this approach would have been provided a range of plausible ICERs if only the all the variables had been taken into account.

(8) Retreatment interval of rituximab

The NICE guidance issued as a result of TA195[25] states on the issue of the retreatment interval of rituximab that "the Committee concluded that an 8.7-month retreatment interval is likely to overestimate the time between consecutive courses of rituximab. However, on the basis of the clinical specialists' advice, the Committee considered that it was unlikely to be as frequent as every 6 months for every person receiving rituximab" (Section 4.3.21). The company assumed a mean retreatment interval of 6 months for its base case analysis. It also included a scenario analysis exploring the impact on the ICER of a 9-month treatment interval, the value assumed to be correct in TA375.[24] This

analysis noticeably increased the ICER from £34,516 per QALY gained to £49,618 per QALY gained.

(9) Appropriateness of assuming treatment duration of TNFis is equal to that of other bDMARDs

The CS[31] assumes within its base case analysis that the time to discontinuation of the first therapy is equal for all bDMARDs. More precisely, it relies on the approach taken by the AG for TA195,[25] adopting the Weibull distribution (and the specific parameter values) used for TNFis in TA195 and applying it to all bDMARDs. However, the report of the AG for TA195 reports different parameter values for the Weibull distribution for RTX and ABA, leading to considerably different treatment duration means: 4.06 years for TNFis, 11.31 years for RTX, and 6.17 years for ABA. The CS[31] also included a scenario analysis exploring the impact of using a different value for the scale parameter of the Weibull distribution for non-TNFi biologics (0.2208 instead of 0.4416). With this alternative value, the result of the analysis for Population A changed from an ICER of £34,516 per QALY gained for the intervention sequence to the intervention sequence being dominated by its comparator. The ERG believes the appropriate parameter values for the respective drugs reported in the AG's report in TA195 should have been used, especially given the importance of its impact. For information, the approach taken in TA375[24] assumed that the time on treatment was dependent on the EULAR response obtained, but independent of the bDMARD that produced this response.

(10) *Appropriateness of assuming changes in HAQ score affect mortality*

Within the CS it is assumed that the mortality rate is linked to changes in the HAQ score. The company based this assumption on the assertion by Norton *et al.*[68] that "HAQ score assessed at 1 year was a significant predictor of mortality after adjustment for baseline clinical and demographic factors". However, the ERG believes this assertion not to be sufficient evidence to assume that changes in HAQ score throughout the disease duration to be predictive of changes in mortality. In addition, a study by Michaud *et al.*[69] concluded that "changes in the PCS [(Physical Component Summary)] and HAQ did not contribute substantially to predictive value over and above the baseline values of these variables". The Assessment Group in TA375 assumed that only baseline HAQ score affects mortality, an assumption also used by the ERG.

(11) Failure to age-adjust utilities

Due to the company's assumption that changes in HAQ score affects mortality, different treatments can result in different simulated mean survival. As such, the ERG believes that the utilities within the model should have been age-adjusted to account for reduced utility in the older population.

(12) Modelling of HAQ improvement in responders for subsequent therapies

The improvement in HAQ scores for subsequent therapies was based on data reported by Emery *et al.*[40] which was estimated from data collected in the RADIATE trial. Emery *et al.*[40] reports that HAQ values changed from baseline by -0.39 for the TOC group compared with -0.05 in control group. These two values were used in the company's economic model to estimate HAQ changes for bDMARDs and cDMARDs respectively. However, given that after six months the non-responders, which are estimated to comprise the 32.3% and 83.5% of patients for bDMARDs and cDMARDs, respectively (also based on TOC and control groups in RADIATE), are assumed to discontinue, those remaining on treatment (i.e., the responders) are likely to experience a higher HAQ improvement than the average of all patients. The company acknowledged the issue (see clarification response,[28] question B5), stating that given that HAQ improvement data were not reported by responder status, they had decided to use full population values instead. The ERG believes that a more reasonable choice is to assume that non-responders would have no HAQ improvement and therefore all the HAQ improvement observed in the full population can be assigned to responders. This implies that responders on bDMARDs would experience an improvement of -0.39/0.677 = -0.576 whilst responders on cDMARDs would experience an improvement of -0.05/0.165=-0.303. The difference in these two figures could potentially be justified by the fact that the proportion of good responders is likely to be considerably higher in patients on bDMARDs than for patients on cDMARDs. In addition, the -0.576 figure for people on bDMARDs is in line with the mean HAQ improvement estimated for responders on a first therapy of TOC + MTX, which is ((-0.322*0.4) +(-0.744*0.489))/ (0.4+0.489) = -0.554, where -0.322 and -0.744 are the mean HAQ improvements for moderate and good EULAR responders, respectively, and 0.4 and 0.488 are the probabilities of moderate and good EULAR response respectively. For information, the approach taken in TA375[24] assumed that the HAQ improvement was dependent on the EULAR response obtained, but not on the treatment that produced this response.

(13) Modelling of treatment discontinuation for subsequent therapies

The company assumed a constant discontinuation rate for subsequent therapies. A constant discontinuation rate was used for subsequent therapies, estimated to be equal to the discontinuation rate between months 6 and 12 of the Weibull distribution used to model time to discontinuation of the first treatment in the model (the second-line bDMARD). The company justified using a constant rate for subsequent therapies instead of a Weibull distribution, as for the first therapy, citing the limitation of cohort models whereby "once a patient enters the "all subsequent period" state their history in terms of time spent on subsequent therapy is lost" (see clarification response,[28] question B5). However, the ERG notes that the hazard of event of a Weibull whose shape parameter is lower than one decreases over time. Therefore, the ERG notes that using the discontinuation rate of the Weibull between 6 and 12 months as a constant rate is likely to underestimate the treatment duration. The

ERG considers that a more appropriate approximation to calculating the discontinuation rate would have been to calculate a constant rate that resulted in the same mean time to discontinuation as that of the Weibull distribution used for the first therapy.

(14) Inaccuracy in TOC (IV) dosing

The ERG notes that the company's economic model does not take the 80 mg dose of TOC (IV) into account and requested clarification from the company on this issue. The company responded that "the 200 mg dose of tocilizumab was selected as the median dose available on the BNF (i.e. 80 mg, 200 mg, 400 mg)" (see clarification response,[28] question B16). The ERG notes that in a drug where the dosing is variable and depends on the weight of the patient, ignoring a smaller dose leads to overestimation of its cost. For example, if a patient required 150mg this could be achieved with two 80 mg vials rather than purchasing one 200 mg vial.

In addition, the ERG noted that the SmPC for TOC states that "for individuals whose body weight is more than 100 kg, doses exceeding 800 mg per infusion are not recommended".[85] Therefore, the ERG believes this limit should have been included when calculating the average vials per dose used in the company's economic model. Failure to include this limit overestimates the cost of TOC (IV) and therefore overestimates the ICER of TOC(IV) compared with CZP.

(15) Approximation of the distribution of weight using weight bands

In order to calculate the average number of vials required for drugs for which the dosing is weight-dependent, the company modelled the weight distribution of the target population. For this purpose, the company divided the range between 40 and 120 kg in 5 kg bands and included an extra band for people between 120 kg and 200 kg. The company then calculated the incidence of each weight band in the TNFi-experienced population of the REALISTIC study.[34] The model calculates the average number of vials required for each drug based in the dosage of the vials and the dose required per kg. The ERG notes potential inaccuracies are introduced because the number of vials needed might differ within a weight band. For example, the dose for IFX is 3 mg/kg and therefore within the 65 to 70 kg band, some will need 2 vials (people whose weight is lower than 66.7) whilst others will need 3. Assigning a distribution to estimate the weight of patients would allow this calculation to be performed more accurately. The current approach is slightly underestimating the dosing of IFX (by assuming patients below 70 kg can have only 200 mg, when patients above 66.7 should have 300 mg) and therefore overestimates the ICER of CZP + MTX compared with IFX+MTX.

(16) Inconsistency in benefits of treatment response during the first cycle

For their base case analysis, the company assumes that the improvement in utility derived from the first treatment is achieved after six weeks of treatment. The ERG notes however, that this approach is

not consistently applied to other variables, such as the HAQ score, and outcomes that depend on the HAQ score, such as mortality and HAQ related costs. This slight inaccuracy is likely to underestimate the benefits of the most efficacious bDMARDs, and therefore overestimate the ICER of CZP compared with ABA and the rest of the TNFis and also overestimate the ICER of TOC compared with CZP. However, the ERG notes that the impact of this inaccuracy on the ICERs is likely to be negligible.

(17) Exclusion of adverse events

Adverse events were excluded from the economic analysis, claiming that there was no meaningful difference in the toxicity or risks of AEs between alternative bDMARDs. However, given that the company's model assumes treatment efficacy affects mortality, AEs could be more predominant in patients that live longer. However, the ERG notes that the impact of excluding AEs from the economic analysis is likely to be negligible based on sensitivity analyses performed in TA375.[24]

(18) Inaccuracies in number of doses per cycle

The ERG notes that there is a slight inconsistency in the number of doses assumed for the first six months of ABA (IV) and IFX therapy compared with the subsequent six-month cycles of these drugs or with drugs whose dosing frequency is not a divisor of the cycle length, such as TOC (IV). Even if for ABA (IV) and IFX 8 and 5 doses are respectively administered during the first six months, only 7.5 and 4.5 should be accounted for in the first cycle. However, the ERG notes that the impact of this inaccuracy on the ICERs is likely to be slight.

(19) Appropriateness of using EQ-5D data from the PREDICT study

The utilities used in the company's model were based on EQ-5D data collected in the PREDICT study. However, the population characteristics used in the model were based on the TNFi experienced population of REALISTIC. The ERG considers that it would have been preferable to use EQ-5D scores collected in REALISTIC and that should there be significant differences between the two populations, the utilities currently used in the model would be biased.

(20) Perceived model errors and other issues surrounding model implementation

The cost of TOC (IV) monotherapy was incorrectly calculated because of an error in the model implementation, which led to erroneous results being reported in the scenario analysis for Population C. The error is that the TOC monotherapy administration costs are linked to those of ABA's method of administration (SC or IV).

The ERG believes that the cohort modelling approach used by the company is not the most appropriate to reflect the nature of the disease and that an individual patient model would be

preferable. Cohort models cannot, as admitted by the company, accurately reflect non-linear functions such as the mortality associated to HAQ score assumed by the company, or the mapping from HAQ score to EQ-5D preferred by the ERG.[65] The ERG notes that in TA375[24] the Assessment Group's model and four of the six models submitted by the companies were individual patient models.

5.4 Additional exploratory analyses undertaken by the ERG

The ERG undertook a number of additional sensitivity analyses using the company's model, in order to address the issues described in the previous section. In this section, the differences between the company's and the ERG's base case analyses are described and the results for of the ERG's base case analyses are reported together with additional scenario analyses using this base case. The ERG notes the PAS currently in place for TOC (IV) and ABA (IV) were not taken into account in these analyses as these are designated as being commercial-in-confidence. The results of the analyses including the PAS are described in the confidential appendix.

5.4.1 Base case analysis

The ERG's base case analysis includes the following amendments, which are described in more detail in Section 5.3.2:

1. Correction of technical programming errors in the company's model.
2. Adding two other sequences to be compared for Population A.
3. Removing ABA treatment from the intervention and comparator sequences for Population A.
4. Using the results of the NMA including J-RAPID.
5. Setting RTX retreatment interval to 7.35. The Appraisal Committee for TA195 concluded that the average retreatment interval was between 6 and 8.7 months.[25] The ERG used the midpoint between these two figures: (6+8.7)/2= 7.35.
6. Using different HAQ improvement for subsequent therapies. Instead of the -0.39 and -0.05 mean change in HAQ score for responders to subsequent bDMARD and cDMARD treatments respectively values of -0.576 for bDMARD responders and -0.303 for cDMARD responders.
7. Using the Weibull parameters reported in TA195[25] for RTX (see Table 56) instead of assuming the same time to discontinuation as for TNF inhibitors.
8. Assume that mortality is only affected by the baseline HAQ score, and that changes in the HAQ score do not affect mortality.
9. Using constant discontinuation rates for subsequent bDMARD treatments that would match the mean treatment duration estimated by the Weibull distribution used for the first treatment line considered in the model (see Table 56).
10. Including the 80 mg dose of TOC (IV) and 800 mg limit for people with a body weight greater than 100 kg.

11. Using amended numbers of administrations per cycle for IFX (3.25) and TOC IV (7 in the first cycle).
12. Including the SC formulations of ABA and TOC, IFX biosimilars and Benepali (a new ETA biosimilar) as comparators in its analyses. Benepali is administered weekly as a 50mg/ml solution for injection in a pre-filled syringe or pre-filled pen. The cost to the NHS of each dose reported in MIMS[86] (in May 2016) is £164.00.

Table 56: Parameters used in the ERG analyses to calculate treatment discontinuation

	First treatment discontinuation (Weibull distribution)				Subsequent treatment discontinuation
	Scale		Shape		Constant annual rate
	Mean	SE	Mean	SE	
CZP	0.4414	0.009584	0.7008	0.034184	0.25
ADA	0.4414	0.009584	0.7008	0.034184	0.25
ETA	0.4414	0.009584	0.7008	0.034184	0.25
GOL	0.4414	0.009584	0.7008	0.034184	0.25
ABA	0.2517	0.006328	0.81	0.038776	0.16
RTX	0.4620	0.034717	0.474	0.036225	0.09
TOC*	0.2517	0.006328	0.81	0.038776	0.16

*Not reported in TA195, assumed to be equal to ABA

The ERG would have preferred to have implemented the mapping from HAQ score to the EQ-5D using the mixture model proposed by Hernández Alava *et al.*[65] unfortunately, this was not possible within the timeline of the STA.

Population A

For Population A, the ERG compared four possible sequences as shown in Table 57.

Table 57: Sequences considered in the ERG's evaluation of Population A

Sequence Number	**1**	**2**	**3**	**4**
Sequence name	CZP before RTX	CZP after RTX	CZP instead of RTX	RTX
First	CZP + MTX	RTX + MTX	CZP + MTX	RTX + MTX
Second	RTX + MTX	CZP + MTX	TOC + MTX	TOC + MTX
Third	TOC(SC) + MTX	TOC(SC) + MTX	M + H + S	M + H + S
Fourth	M + H + S	M + H + S	NBT	NBT
Fifth	NBT	NBT	Palliative care	Palliative care
Sixth	Palliative care	Palliative care		

NBT = Non-biologic treatment: a weighted mix of leflunomide, gold, ciclosporin, azathioprine (25% each)
M + H + S = MTX + HCQ + SSZ

In order to evaluate Sequence 2, the ERG had to adapt the company's model to support CZP + MTX as a follow-up treatment. In order to do so, the ERG calculated the probability of no response for CZP + MTX following the same approach as the company used to calculate the probability of no response for ABA + MTX and RTX + MTX: using the treatment effect parameters in the NMA to the trial-specific baseline effects from the RADIATE study. The resulting probability of discontinuation for CZP + MTX was estimated to be 44.6%. The ERG used the SC formulation of TOC instead of IV within the sequences because it was less expensive and was assumed to have the same efficacy.[87]

The results of the deterministic and probabilistic analyses using the ERG base case are shown in Table 58 and Table 59. The ERG notes that CZP after RTX (Sequence 2) dominates CZP before RTX (Sequence 1) and that the currently recommended pathway (Sequence 4) dominates the same sequence if RTX is replaced with CZP (Sequence 4). The ERG notes that these results [redacted]

[redacted]

Figure 25 shows the cost-effectiveness acceptability curve (CEAC) for the ERG's base case analysis for Population A. The ERG notes that given the limitations of the model when comparing elongated sequences, the value of the CEAC is limited.

It is noticeable that Sequence 4, which includes only three lines of biologics, dominates Sequences 1 and 3, both of which include four lines of biologics. The ERG believes that this is due to the different methods used for modelling first treatments compared with follow-up treatments, in particular, the fact that the benefits of the first treatment outweigh those of subsequent treatments. For this reason, the ERG believes the model is not appropriate for making comparisons of sequences which include different numbers of treatments and that the fully incremental analyses reported in Table 58 and Table 59 should be interpreted with caution. However, the conclusion that CZP should not be placed before RTX appears to be robust.

Table 58: Results of the ERG's base case analysis for Population A (deterministic)

Sequences		**Total QALYs**	**Total costs**	**Inc. QALYs**	**Inc. costs**	**ICER (£/QALY)**
3	CZP instead of RTX‡	7.719	£125,364	-	-	Dominated
1	CZP before RTX‡	8.239	£133,780	-	-	Dominated
4	RTX‡	8.378	£122,451	-	-	-
2	CZP after RTX‡	8.649	£130,016	0.271	£7,565	**£27,946**

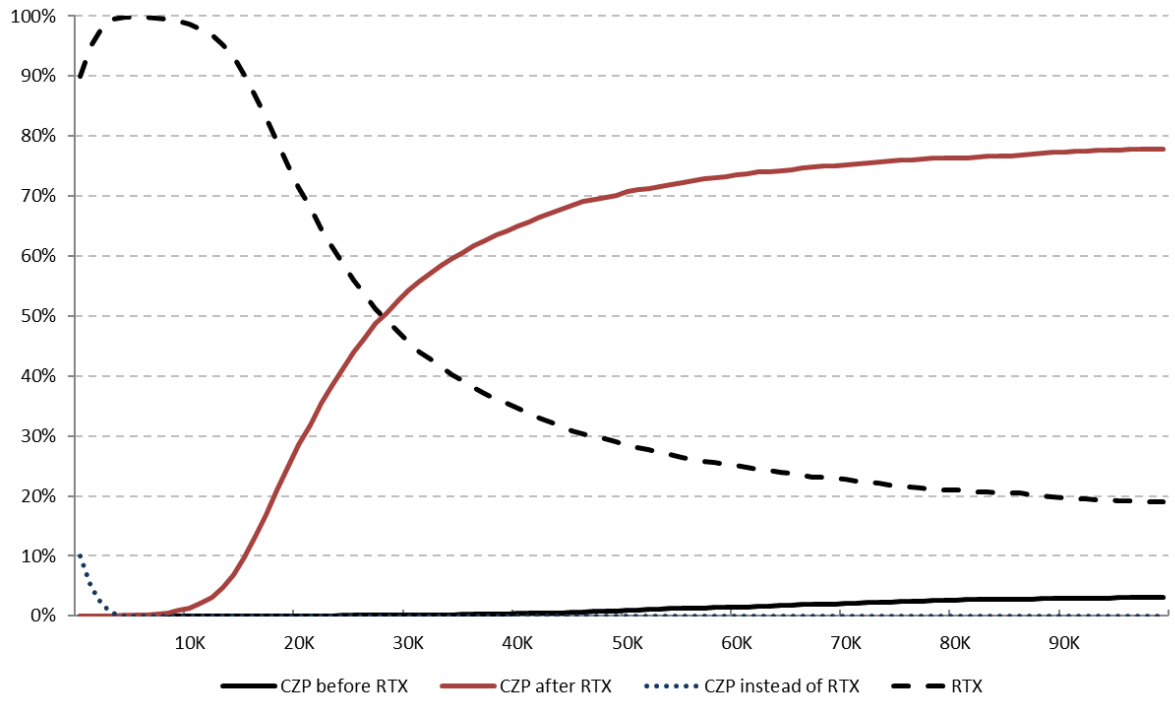
†Rest of the sequence: TOC(SC)+MTX, MTX + HCQ + SSZ, NBT, PC
‡CiC PAS not included

Table 59: Results of the ERG's base case analysis for Population A (probabilistic)

Sequences†		**Total QALYs**	**Total costs**	**Inc. QALYs**	**Inc. costs**	**ICER (£/QALY)**	**Probability of cost-effectiveness at a threshold of**	
							£20,000 /QALY	**£30,000 /QALY**
3	CZP instead of RTX‡	7.796	£128,376	-	-	Dominated	0.0000	0.0000
1	CZP before RTX‡	8.347	£136,751	-	-	Dominated	0.0000	0.0020
4	RTX‡	8.461	£125,189	-	-	-	0.7146	0.4564
2	CZP after RTX‡	8.732	£132,692	0.271	£7,504	**£27,700**	0.2852	0.5426

†Rest of the sequence: TOC(SC)+MTX, MTX + HCQ + SSZ, NBT, PC
‡CiC PAS not included

Figure 25: Cost effectiveness acceptability curve for Population A (ERG's base case analysis)



Population B

For Population B, the ERG evaluated the same sequences as the company (previously summarised in Table 28). However, the ERG included additional comparators: the biosimilars for IFX and ETA and the SC formulations of TOC and ABA. The results for the deterministic analyses are presented in Table 60 and the results for the probabilistic analyses are presented in Table 61. The values in Table 61 are subject to Monte Carlo sampling errors which explains the difference between two interventions with assumed identical efficacy. As such, caution should be undertaken in interpreting results where the difference in QALYs between strategies are very small. Figure 26 shows the CEAC for the ERG's base case analysis for Population B.

Table 60: Results of the ERG's base case analysis for Population B (deterministic)

First therapy of the sequence†	Total QALYs	Total costs	Inc. QALYs	Inc. costs	ICER (£/QALY)
ETA(bio) + MTX	6.897	£91,635	-	-	-
IFX(bio) + MTX	6.897	£93,402	-	-	Dominated
ADA + MTX	6.897	£93,524	-	-	Dominated
GOL + MTX	6.897	£93,524	-	-	Dominated
ETA + MTX	6.897	£93,946	-	-	Dominated
IFX + MTX	6.897	£96,040	-	-	Dominated
CZP + MTX	7.176	£95,197	0.279	£3,562	**£12,773**
ABA(IV) + MTX‡	7.237	£121,272	-	-	Dominated
ABA(SC) + MTX‡	7.237	£125,187	-	-	Dominated
TOC(SC) + MTX‡	7.697	£118,338	0.520	£23,141	**£44,479**
TOC(IV) + MTX‡	7.697	£127,749	-	-	Dominated

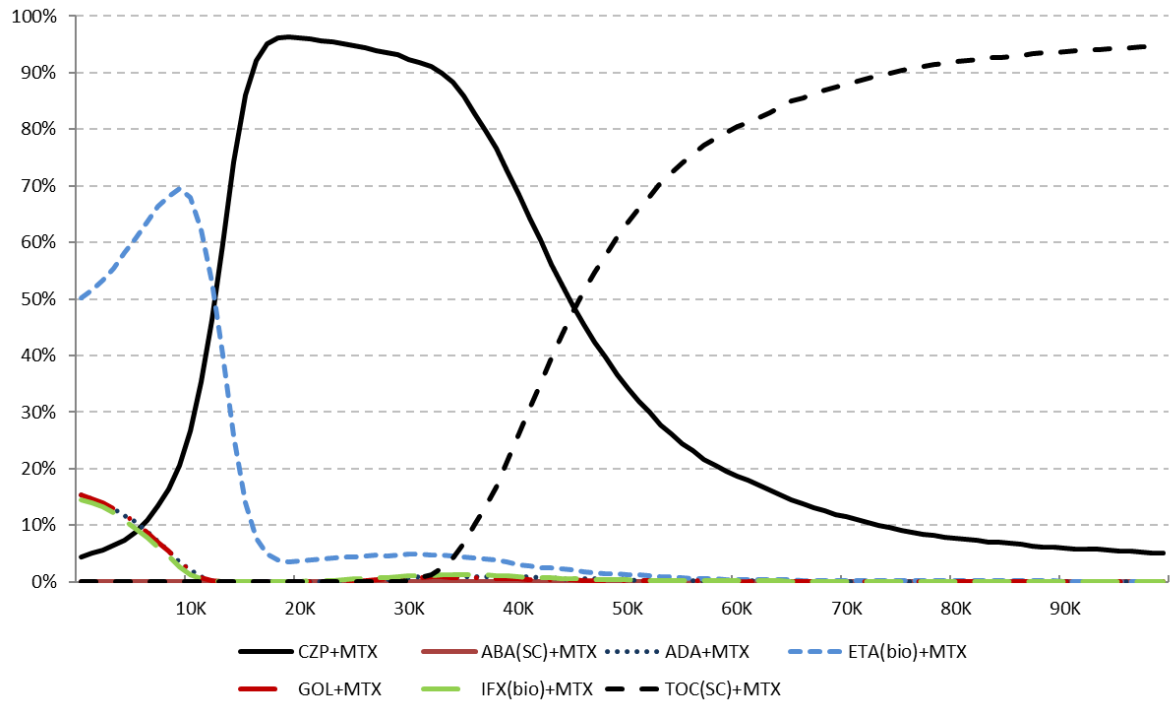
†Rest of the sequence: MTX + HCQ + SSZ, LEF, GLD, CIC, AZA, PC
‡CiC PAS not included
bio = biosimilar

Table 61: Results of the ERG's base case analysis for Population B (probabilistic)

First therapy of the sequence†	Total QALYs	Total costs	Inc. QALYs	Inc. costs	ICER (£/QALY)	Probability of cost effectiveness at a threshold of	
						£20,000 /QALY	£30,000 /QALY
IFX(bio) + MTX	6.932	£94,268	-	-	Dominated	0.0010	0.0106
IFX + MTX	6.933	£96,797	-	-	Dominated	0.0000	0.0000
ETA(bio) + MTX	6.933	£92,507	-	-	-	0.0368	0.0486
ADA + MTX	6.933	£94,280	-	-	Extendedly dominated	0.0000	0.0024
ETA + MTX	6.935	£94,724	-	-	Dominated	0.0006	0.0044
GOL + MTX	6.936	£94,334	-	-	Extendedly dominated	0.0000	0.0018
CZP + MTX	7.213	£95,899	0.280	£3,392	**£12,116**	0.9622	0.9230
ABA(SC) + MTX‡	7.271	£126,084	-	-	Dominated	0.0000	0.0000
ABA(IV) + MTX‡	7.274	£122,109	-	-	Dominated	0.0000	0.0000
TOC(SC) + MTX‡	7.725	£119,171	0.512	£23,272	**£45,414**	0.0000	0.0028
TOC(IV) + MTX‡	7.725	£128,417	-	-	Dominated	0.0000	0.0000

†Rest of the sequence: MTX + HCQ + SSZ, LEF, GLD, CIC, AZA, PC
‡CiC PAS not included
bio = biosimilar

Figure 26: Cost effectiveness acceptability curve for Population B (ERG's base case analysis)



Population C

For Population C, the ERG evaluated the same sequences as the company (previously summarised in Table 29). In addition, the ERG included the biosimilar for ETA and the SC formulation of TOC. The results for the deterministic and probabilistic analyses are presented in Table 62 and Table 63. Figure 27 shows the CEAC for the ERG's base case analysis for Population C.

Table 62: Results of the ERG's base case analysis for Population C (deterministic)

First therapy of the sequence†	Total QALYs	Total costs	Inc. QALYs	Inc. costs	ICER (£/QALY)
ETA(bio)	6.745	£89,853	-	-	-
ADA	6.745	£91,657	-	-	Dominated
ETA	6.745	£92,058	-	-	Dominated
CZP	7.024	£93,807	0.279	£3,953	**£14,185**
TOC(SC) ‡	7.528	£117,033	0.505	£23,226	**£46,018**
TOC(IV) ‡	7.528	£126,262	-	-	Dominated

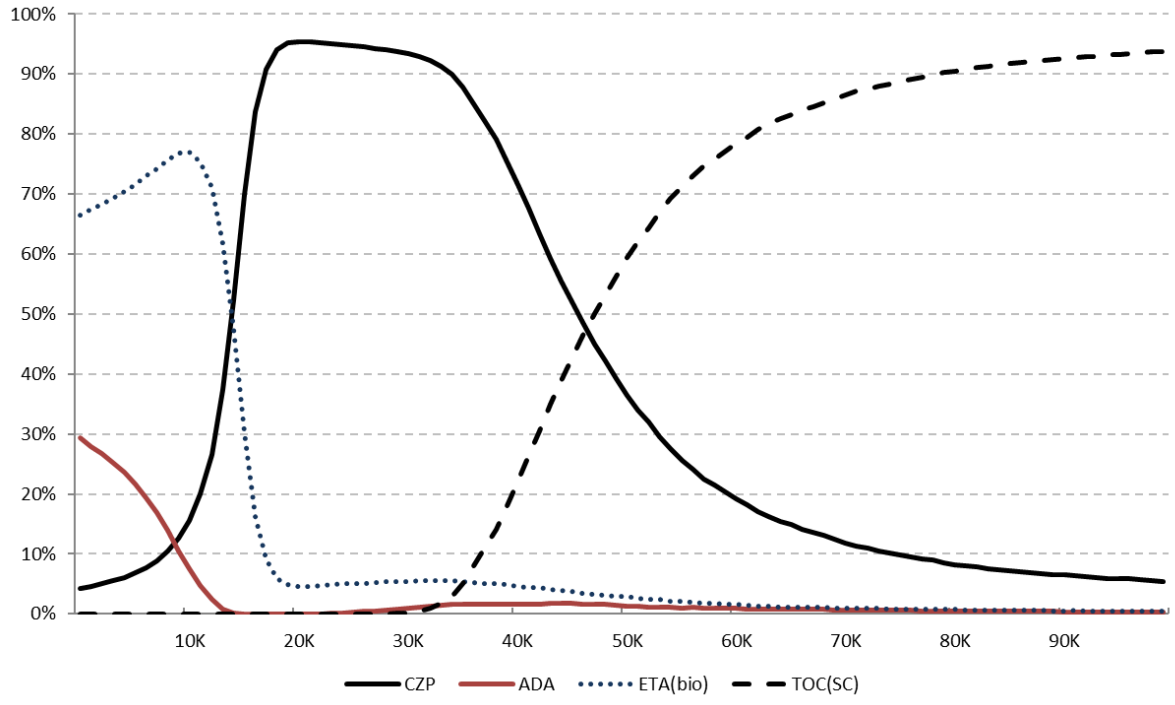
† Rest of the sequence: LEF, GLD, CIC, AZA, PC
‡ CiC PAS not included
bio = biosimilar

Table 63: Results of the ERG's base case analysis for Population C (probabilistic)

First therapy of the sequence†	Total QALYs	Total costs	Inc. QALYs	Inc. costs	ICER (£/QALY)	Probability of cost effectiveness at a threshold of	
						£20,000 /QALY	£30,000 /QALY
ETA(bio)	6.781	£90,323	-	-	-	0.0464	0.0546
ADA	6.786	£92,121	-	-	Extendedly dominated	0.0000	0.0090
ETA	6.787	£92,534	-	-	Extendedly dominated	0.0000	0.0000
CZP	7.070	£94,311	0.289	£3,988	**£13,784**	0.9536	0.9348
TOC(SC) ‡	7.561	£117,142	0.491	£22,832	**£46,501**	0.0000	0.0016
TOC(IV) ‡	7.566	£126,323	0.005	£9,181	**£1,945,969**	0.0000	0.0000

† Rest of the sequence: LEF, GLD, CIC, AZA, PC
‡ CiC PAS not included
bio = biosimilar

Figure 27: Cost effectiveness acceptability curve for Population C (ERG's base case analysis)



5.4.2 Scenario analyses

Assuming other TNFis have the same efficacy as CZP

The company's systematic review failed to identify studies reporting the efficacy of IFX, ETA and ADA in combination with MTX in TNFi-IR patients. In the absence of evidence, the company assumed their efficacy to be equal to that of GOL + MTX, on the basis that they were all TNFis. Similarly, the company assumed ADA and ETA monotherapies to be equally effective as GOL in

TNFi-IR patients. The ERG notes that CZP is also a TNFi and that following the same reasoning, it would be equally valid to assume that IFX, ETA and ADA in combination with MTX are as effective as CZP + MTX and that ETA and ADA monotherapies are as effective as CZP monotherapy in TNFi-IR patients. Table 64 and Table 65 present the results of the analyses with this alternative assumption only for Populations B and C given that Population A remains unaffected. The results show that CZP + MTX is dominated by the less expensive ETA biosimilar + MTX.

Table 64: Results of the ERG's scenario analysis assuming other TNFis are as effective as CZP for Population B (deterministic)

First therapy of the sequence†	Total QALYs	Total costs	Inc. QALYs	Inc. costs	ICER (£/QALY)
GOL + MTX	6.897	£93,524	-	-	-
ETA(bio) + MTX	7.176	£94,943	0.279	£1,418	**£5,085**
CZP + MTX	7.176	£95,197	-	-	Dominated
IFX(bio) + MTX	7.176	£96,619	-	-	Dominated
ADA + MTX	7.176	£97,193	-	-	Dominated
ETA + MTX	7.176	£97,694	-	-	Dominated
IFX + MTX	7.176	£99,719	-	-	Dominated
ABA(IV) + MTX‡	7.237	£121,272	-	-	Dominated
ABA(SC) + MTX‡	7.237	£125,187	-	-	Dominated
TOC(SC) + MTX‡	7.697	£118,338	0.520	£23,395	**£44,967**
TOC(IV) + MTX‡	7.697	£127,749	-	-	Dominated

†Rest of the sequence: MTX + HCQ + SSZ, LEF, GLD, CIC, AZA, PC
‡CiC PAS not included
bio = biosimilar

Table 65: Results of the ERG's scenario analysis assuming other TNFis are as effective as CZP for Population C (deterministic)

First therapy of the sequence†	Total QALYs	Total costs	Inc. QALYs	Inc. costs	ICER (£/QALY)
ETA(bio)	7.024	£93,629	-	-	-
CZP	7.024	£93,807	-	-	Dominated
ADA	7.024	£95,816	-	-	Dominated
ETA	7.024	£96,304	-	-	Dominated
TOC(SC) ‡	7.528	£117,033	0.505	£23,404	**£46,371**
TOC(IV) ‡	7.528	£126,262	-	-	Dominated

† Rest of the sequence: LEF, GLD, CIC, AZA, PC
‡ CiC PAS not included
bio = biosimilar

NMA excluding J-RAPID

The company excluded J-RAPID from the NMA for their base case analysis "because of small sample size and the associated risk of a biased effect size estimate."[31] The ERG consider that the J-RAPID study should only be excluded if there is evidence to suggest that the study was of poor quality and therefore included it in the NMA for its base case analysis. However, given that J-RAPID was conducted solely in Japan and that patients had prior low MTX use (6-8 mg/week in line with licensed dose in Japan at time of trial) the ERG decided to undertake a scenario analysis excluding J-RAPID. The results show that the impact of this alternative assumption in the ICER are limited.

Table 66: Results of the ERG's scenario analysis excluding J-RAPID from the NMA for Population A (deterministic)

Sequences		Total QALYs	Total costs	Inc. QALYs	Inc. costs	ICER (£/QALY)
3	CZP instead of RTX	7.671	£124,990	-	-	Dominated
1	CZP before RTX	8.195	£133,437	-	-	Dominated
4	RTX	8.378	£122,451	-	-	-
2	CZP after RTX	8.633	£129,441	0.255	£6,990	**£27,406**

Table 67: Results of the ERG's scenario analysis excluding J-RAPID from the NMA for Population B (deterministic)

First therapy of the sequence†	Total QALYs	Total costs	Inc. QALYs	Inc. costs	ICER (£/QALY)
ETA(bio) + MTX	6.897	£91,635	-	-	-
IFX(bio) + MTX	6.897	£93,402	-	-	Dominated
ETA + MTX	6.897	£93,524	-	-	Dominated
ADA + MTX	6.897	£93,524	-	-	Dominated
GOL + MTX	6.897	£93,946	-	-	Dominated
IFX + MTX	6.897	£96,040	-	-	Dominated
CZP + MTX	7.123	£94,468	0.226	£2,833	**£12,531**
ABA(IV) + MTX‡	7.237	£121,272	-	-	Dominated
ABA(SC) + MTX‡	7.237	£125,187	-	-	Dominated
TOC(SC) + MTX‡	7.697	£118,338	0.573	£23,870	**£41,654**
TOC(IV) + MTX‡	7.697	£127,749	-	-	Dominated

†Rest of the sequence: MTX + HCQ + SSZ, LEF, GLD, CIC, AZA, PC
‡CiC PAS not included
bio = biosimilar

Table 68: Results of the ERG's scenario analysis excluding J-RAPID from the NMA for Population C (deterministic)

First therapy of the sequence†	Total QALYs	Total costs	Inc. QALYs	Inc. costs	ICER (£/QALY)
ETA(bio)	6.745	£89,853	-	-	-
ADA	6.745	£91,657	-	-	Dominated
ETA	6.745	£92,058	-	-	Dominated
CZP	6.970	£93,030	0.225	£3,177	**£14,113**
TOC(SC) ‡	7.528	£117,033	0.558	£24,003	**£42,989**
TOC(IV) ‡	7.528	£126,262	-	-	Dominated

† Rest of the sequence: LEF, GLD, CIC, AZA, PC
‡ CiC PAS not included
bio = biosimilar

5.5 Discussion

The CS includes a systematic review of economic evaluations of treatments for moderate to severe RA in TNFi-IR patients together with a *de novo* model-based economic evaluation of CZP versus currently recommended treatments in adult TNFi-IR patients.

The company's systematic review of existing economic evaluations did not identify any studies that estimated the cost effectiveness of CZP in patients who had failed on a previous TNFi.

The company's *de novo* economic model adopts a Markov cohort approach to estimate costs and health outcomes for CZP and its comparators from the perspective of the NHS and PSS over a lifetime horizon. The analyses presented in the CS relate to three different subpopulations of patients who have had an inadequate response to a TNF inhibitor: Population A: patients eligible for RTX + MTX; Population B: patients for whom RTX is contraindicated or withdrawn, and; Population C: adults for whom RTX therapy cannot be given because MTX is contraindicated or withdrawn. The characteristics of the population modelled are based on the characteristics of the TNFi-IR population in REALISTIC. The model uses a six-month cycle length. For each treatment, patients are assumed to go through a six-month treatment period and then discontinue treatment unless they achieve a good or moderate EULAR response. For the first treatment considered in the model (the second line bDMARD), EULAR response probabilities are modelled using the results of an NMA. The NMA includes results of trials for CZP, TOC, ABA, RTX and GOL in combination with MTX. In the absence of data, the efficacy of ETA, ADA and IFX is assumed to be equal to GOL (TNFi class equivalence). For follow up treatments, the probabilities of EULAR response were derived from the results of RADIATE; the results of TOC + MTX are extrapolated to other bDMARDs and those of MTX to cDMARDs. Patients also discontinue follow-up treatments at six-months if no EULAR response is observed. Patients who discontinue treatment are assumed to start the next treatment in the

sequence immediately. Health utilities are modelled according to HAQ score progression. HAQ score improves following a positive response to treatment, and after response the HAQ score is assumed to stay constant for the duration of bDMARD treatment, but contrastingly is assumed to increase linearly for patients on cDMARDs until a maximum value is reached. Resource use estimates were based on previous NICE TAs and the views of an expert clinician. Unit costs were taken from the BNF, the PSSRU and NHS Reference Costs 2014 to 2015. AEs are assumed not to have an impact on the relative HRQoL and costs.

Based on the probabilistic version of the company's base case model, adding a CZP + MTX treatment before the NICE recommended treatment sequence in Population A is expected to produce an additional 0.290 QALYs at an additional cost of £9,647; the probabilistic ICER for the sequence including CZP + MTX versus the currently recommended treatment sequence is expected to be £33,222 per QALY gained.

For Population B, according to the company's probabilistic analysis, CZP + MTX is expected to produce 0.256 additional QALYs at an additional cost of £884 compared with GOL + MTX, resulting in an ICER of £3,461 per QALY gained. TOC + MTX is expected to produce 0.201 additional QALYs at an additional cost of £26,659 compared with CZP + MTX, resulting in an ICER of £132,783 per QALY gained.

For Population C, based on the company's probabilistic analysis, CZP monotherapy is expected to produce an additional 0.260 QALYs at an additional cost of £1,336 compared with ADA, resulting in an ICER of £5,151 per QALY gained. TOC(IV) is expected to produce an additional 0.196 QALYs at an additional cost of £26,179 compared with CZP monotherapy, resulting in an ICER of £133,655 per QALY gained.

The ERG notes that the results for Populations A, B and C do not include: biosimilars of IFX and ETA, SC formulations of TOC and ABA, or the commercial-in-confidence PAS in place for TOC and ABA.

The ERG's critical appraisal identified a number of issues relating to the company's model and analysis.

The ERG undertook a series of exploratory analyses based on the company's submitted model. First, the ERG produced their base case analysis applying the following changes to the company's base case: (i) correction of technical programming errors in the company's model; (ii) adding two sequences to be compared in Population A; (iii) removing ABA + MTX treatment after TOC + MTX

from the compared sequences in Population A; (iv) using the results of the NMA including J-RAPID; (v) setting the RTX retreatment interval to 7.35; (vi) adjusting mean HAQ improvements reported in RADIATE to responders; (vii) using different parameters for the Weibull distribution modelling time to treatment discontinuation for different treatments as reported in TA195;[25] (viii) assuming only baseline HAQ score affects mortality, (ix) using constant discontinuation rates for subsequent treatment that match the mean of the Weibull distribution used for the first treatment; (x) considering the 80 mg vial of TOC and applying a 800mg dose limit per administration of TOC as recommended in its SmPC;[85] (xi) using amended administrations per cycle for TOC and IFX, and; (xii) adding biosimilars of IFX and ETA and SC formulations of TOC and ABA as comparators. Two additional scenario analyses were undertaken using this base case to explore the impact of: (i) assuming ADA, IFX and ETA have the same efficacy as CZP (instead of GOL); and (ii) excluding J-RAPID from the NMA, as was done within the company's base case.

For Population A, the ERG added two sequences to those compared by the company; in addition to comparing the currently recommended sequence with a sequence where CZP + MTX treatment was prepended, the ERG considered sequences in which CZP + MTX was provided after the RTX + MTX treatment and where CZP + MTX was provided instead of RTX + MTX. The results of the ERG's exploratory analyses show that: (i) the sequence where CZP + MTX is provided before RTX + MTX is dominated by the sequence where RTX + MTX is used before CZP + MTX; and, (ii) the sequence where CZP + MTX is provided instead of RTX + MTX is dominated by the currently recommended sequence. The ERG notes that these results suggest that on the basis of the model CZP + MTX should not be given before or instead of RTX + MTX [redacted]

[redacted]

For Population B, ERG's base case analysis shows comparable results to those presented within the CS, but the range of WTP for which CZP + MTX is the most cost-effective option is noticeably narrower: the probabilistic ICER for CZP + MTX compared with ETA(bio) + MTX is £12,116 per QALY gained whilst the probabilistic ICER of TOC(SC) + MTX versus CZP + MTX is £45,414.

Similarly, for Population C, the ERG's base case analysis produces comparable results to those reported by the company. However, the probabilistic ICER of CZP monotherapy versus ETA(bio) is £13,784 per QALY gained and the probabilistic ICER of TOC(SC) monotherapy compared with CZP monotherapy is £46,501 per QALY gained. The range of WTP for which CZP monotherapy is the most cost-effective option is noticeably narrower.

The ERG notes that the PAS currently in place for TOC (IV) and ABA (IV) were not taken into account for these analyses. The results of the analyses including the PAS are described in the confidential appendix.

There remain several potentially important areas of uncertainty:

1. The lack of data on the efficacy of ETA, ADA and IFX in combination with MTX in TNFi-IR patients. Similarly, the lack of data on the efficacy of ETA and ADA monotherapy in TNFi-IR patients. If ETA is assumed to be as efficacious as CZP then CZP would be dominated by the biosimilar for ETA.
2. The scarcity of data on the efficacy of bDMARDs in general and TNFis in particular in patients who have had an inadequate response to two or more bDMARDs.
3. The NMA is considered to be subject to limitations: EULAR no response, moderate response and good response rates should have all come from the REALISTIC,[34] instead of only the no response rate; draws from the joint posterior distribution (i.e. CODA) should have been used instead of using univariate normal distributions or multivariate normal distributions; and, a plausible random effects model should have been used instead of a fixed effects model. As such, the relative effectiveness of the bDMARDs is uncertain, along with the conclusions formed from analyses using the NMA data.

6 END OF LIFE

NICE end of life supplementary advice should be applied in the following circumstances and when both the criteria referred to below are satisfied:

- The treatment is indicated for patients with a short life expectancy, normally less than 24 months and;
- There is sufficient evidence to indicate that the treatment offers an extension to life, normally of at least an additional 3 months, compared to current NHS treatment, and;

The company state that 'CZP for the treatment of RA patients with prior TNFi exposure should not be considered as an end of life treatment' (page 161 of the CS). The ERG would concur with this view and believe that neither criterion would be met as patients receiving treatment would be expected to have a life expectancy considerable more than 24 months and there is little robust evidence to suggest that CZP would provide an additional 3 months of life compared with the comparator bDMARDs.

7 OVERALL CONCLUSIONS

Based on the probabilistic version of the company's base case model, prepending a CZP + MTX treatment to the currently recommended treatment sequence for people eligible for treatment with RTX + MTX (Population A) is expected to produce an additional 0.290 QALYs at an additional cost of £9,647 resulting in an ICER of £33,222 per QALY gained. The company's PSA estimates that there is a probability of 0.37 that prepending a CZP + MTX treatment to the currently recommended sequence has an ICER below £30,000 per QALY gained. For patients for whom RTX is contraindicated or withdrawn (Population B), the estimated ICER of CZP + MTX versus GOL + MTX is £3,527 per QALY gained and the estimated ICER of TOC+ MTX versus CZP + MTX is £132,783 per QALY gained based the company's probabilistic analysis. Assuming a WTP threshold of £30,000 per QALY gained, the probability that CZP + MTX produces more net benefit than its comparators is approximately 0.96. Finally, for patients for whom MTX is contraindicated or withdrawn (Population C), the estimated ICER of CZP monotherapy versus ADA monotherapy is £5,151 per QALY gained and the estimated ICER of TOC monotherapy versus CZP monotherapy is £133,655 per QALY gained based on the company's probabilistic analysis. Assuming a WTP threshold of £30,000 per QALY gained, the probability that CZP + MTX produces more net benefit than its comparators is approximately 0.97.

The ERG's critical appraisal identified a number of issues relating to the company's model and analysis. The most pertinent of these relate to: (i) the weaknesses of the NMA (ii) inclusion of two lines of bDMARDs after RTX + MTX; (iii) exclusion from the base case of IFX and ETA biosimilars; (iv) exclusion from the base case of SC formulations of TOC and ABA; (v) assuming the same treatment duration for all bDMARDs; (vi) assuming a retreatment interval of RTX that was deemed too short by the Committee in TA195;[25] (vii) ignoring the 80 mg dose of TOC and the 800mg limit per administration; (viii) assuming that the mean HAQ improvements reported in RADIATE apply to responders. The ERG undertook a series of exploratory analyses based on the company's submitted model in order to address the limitations listed above, however, no additional work was undertaken correcting the NMA and as such, the level of uncertainty in all presented results is not clear.

The ERG's base case analysis suggests that for Population A CZP + MTX should not be used before RTX + MTX. Limitations of the company's model in the methods for modelling subsequent treatments mean that the fully incremental analysis was not deemed plausible. However, where the length of the sequence was similar the use of RTX + MTX first, or the use of RTX + MTX rather than CZP + MTX was dominant. This result is not unexpected given the similar efficacies of RTX + MTX and CZP and MTX and the lower acquisition price associated with RTX.

For Population B, the probabilistic ICER of CZP + MTX versus ETA(bio) + MTX is expected to be £12,116 per QALY gained and the probabilistic ICER of TOC(SC) + MTX versus CZP + MTX is expected to be £45,414 per QALY gained. These ICERs are less favourable to CZP + MTX than the company's base case ICERs. However, the probability that CZP + MTX produces more net benefit than its comparators assuming a WTP threshold of £30,000 per QALY gained remains essentially unchanged at 0.96.

For Population C, the probabilistic ICER of CZP + MTX versus ETA(bio) monotherapy is expected to be £13,784 per QALY gained and the probabilistic of TOC(SC) monotherapy versus CZP monotherapy is expected to be £46,501 per QALY gained. These ICERs are less favourable to CZP monotherapy than the company's base case ICERs. However, the probability that CZP monotherapy produces more net benefit than its comparators assuming a WTP threshold of £30,000 per QALY gained is reduced slightly to 0.96.

Additional analyses undertaken by the ERG using this revised base case model indicate that: excluding J-RAPID for the NMA has little impact on the results of the analyses. In contrast, assuming that ADA, IFX and ETA in combination with MTX have the same efficacy as CZP + MTX (rather than GOL + MTX) leads to ETA biosimilar + MTX dominating CZP + MTX; similarly, assuming ADA and ETA monotherapy have the same efficacy as CZP monotherapy leads to ETA biosimilar monotherapy dominating CZP monotherapy. The ERG notes that even were CZP + MTX dominated by ETA biosimilar + MTX there remains comparators for which it is estimated that CZP + MTX is dominant, such as IFX + MTX and ADA + MTX. As these will remain options for treatment in Populations B and C a positive recommendation for CZP + MTX would allow an alternative if ETA biosimilar was not appropriate.

With respect to the company's economic analysis and the ERG's additional exploratory analyses, there remain several potentially important areas of uncertainty:

1. The lack of data on the efficacy of ETA, ADA and IFX in combination with MTX in TNFi-IR patients. There is a similar lack of data on the efficacy of ETA and ADA monotherapy in TNFi-IR patients. Different assumptions for the efficacy of these drugs produced markedly different results.
2. The scarcity of data on the efficacy of bDMARDs in general, and TNFis in particular, in patients who have had an inadequate response to two or more bDMARDs. There is also the possibility that there could be reduced efficacy of TNFis following inadequate response to a previous TNFi.

3. The relative efficacies of the bDMARDs are uncertain given the limitations of the NMA within the CS.

[redacted]

8 REFERENCES

1. Scott DL, Steer S. The course of established rheumatoid arthritis. *Best Practice & Research Clinical Rheumatology* 2007;**21**:943-67. http://dx.doi.org/http://dx.doi.org/10.1016/j.berh.2007.05.006
2. Pincus T, Fuchs HA, Callahan LF, Nance EP, Kaye JJ. Early radiographic joint space narrowing and erosion and later malalignment in rheumatoid arthritis: a longitudinal analysis. *J Rheumatol* 1998;**25**:636-40. http://europepmc.org/abstract/MED/9558162
3. Drossaers-Bakker KW, Kroon HM, Zwinderman AH, Breedveld FC, Hazes JMW. Radiographic damage of large joints in long-term rheumatoid arthritis and its relation to function. *Rheumatology* 2000;**39**:998-1003. http://dx.doi.org/10.1093/rheumatology/39.9.998
4. Allaire S, Wolfe F, Niu J, LaValley MP, Zhang B, Reisine S. Current risk factors for work disability associated with rheumatoid arthritis: Recent data from a US national cohort. *Arthritis Care & Research* 2009;**61**:321-8. http://dx.doi.org/10.1002/art.24281
5. Naz SM, Symmons DPM. Mortality in established rheumatoid arthritis. *Best Practice & Research Clinical Rheumatology* 2007;**21**:871-83. http://dx.doi.org/http://dx.doi.org/10.1016/j.berh.2007.05.003
6. Dadoun S, Zeboulon-Ktorza N, Combescure C, Elhai M, Rozenberg S, Gossec L, *et al.* Mortality in rheumatoid arthritis over the last fifty years: Systematic review and meta-analysis. *Joint Bone Spine* 2013;**80**:29-33. http://dx.doi.org/http://dx.doi.org/10.1016/j.jbspin.2012.02.005
7. Meune C, Touzé E, Trinquart L, Allanore Y. Trends in cardiovascular mortality in patients with rheumatoid arthritis over 50 years: a systematic review and meta-analysis of cohort studies. *Rheumatology* 2009;**48**:1309-13. http://dx.doi.org/10.1093/rheumatology/kep252
8. The National Audit Office. *Services for people with rheumatoid arthritis*. 2009. URL: https://www.nao.org.uk/wp-content/uploads/2009/07/0809823.pdf (Accessed 24 March, 2016).
9. Symmons DPM, Barrett EM, Bankhead CR, Scott DGL, Silman AJ. The incidence of rheumatoid arthritis in the United Kingdom: results from the Norfolk Arthritis Register. *Rheumatology* 1994;**33**:735-9. http://dx.doi.org/10.1093/rheumatology/33.8.735
10. Alamanos Y, Drosos AA. Epidemiology of adult rheumatoid arthritis. *Autoimmunity Reviews* 2005;**4**:130-6. http://dx.doi.org/http://dx.doi.org/10.1016/j.autrev.2004.09.002
11. MacGregor AJ, Snieder H, Rigby AS, Koskenvuo M, Kaprio J, Aho K, *et al.* Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins. *Arthritis & Rheumatism* 2000;**43**:30-7. http://dx.doi.org/10.1002/1529-0131(200001)43:1<30::AID-ANR5>3.0.CO;2-B
12. Jones MA, Silman AJ, Whiting S, Barrett EM, Symmons DP. Occurrence of rheumatoid arthritis is not increased in the first degree relatives of a population based inception cohort of inflammatory polyarthritis. *Annals of the Rheumatic Diseases* 1996;**55**:89-93. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1010100/
13. National Collaborating Centre for Chronic Conditions. *Rheumatoid arthritis: national clinical guideline for management and treatment in adults*: Royal College of Physicians (UK); 2009.
14. Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS, *et al.* The american rheumatism association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis & Rheumatism* 1988;**31**:315-24. http://dx.doi.org/10.1002/art.1780310302
15. Combe B, Landewe R, Lukas C, Bolosiu HD, Breedveld F, Dougados M, *et al.* EULAR recommendations for the management of early arthritis: report of a task force of the European Standing Committee for International Clinical Studies Including Therapeutics (ESCISIT). *Annals of the Rheumatic Diseases* 2007;**66**:34-45. http://dx.doi.org/10.1136/ard.2005.044354
16. Aletaha D, Neogi T, Silman AJ, Funovits J, Felson DT, Bingham CO, *et al.* 2010 Rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. *Annals of the Rheumatic Diseases* 2010;**69**:1580-8. http://dx.doi.org/10.1136/ard.2010.138461

17. Felson DT, Anderson JJ, Boers M, Bombardier C, Furst D, Goldsmith C, *et al.* American college of rheumatology preliminary definition of improvement in rheumatoid arthritis. *Arthritis & Rheumatism* 1995;**38**:727-35. http://dx.doi.org/10.1002/art.1780380602
18. van Gestel AM, Prevoo MLL, van't Hof MA, van Rijswijk MH, van de Putte LBA, van Riel PL. Development and validation of the european league against rheumatism response criteria for rheumatoid arthritis: Comparison with the preliminary american college of rheumatology and the world health organization/international league against rheumatism criteria. *Arthritis & Rheumatism* 1996;**39**:34-40. http://dx.doi.org/10.1002/art.1780390105
19. Felson DT. Assessing the efficacy and safety of rheumatic disease treatments: Obstacles and proposed solutions. *Arthritis & Rheumatism* 2003;**48**:1781-7. http://dx.doi.org/10.1002/art.11087
20. Prevoo MLL, Van'T Hof MA, Kuper HH, Van Leeuwen MA, Van De Putte LBA, Van Riel PLCM. Modified disease activity scores that include twenty-eight-joint counts development and validation in a prospective longitudinal study of patients with rheumatoid arthritis. *Arthritis & Rheumatism* 1995;**38**:44-8. http://dx.doi.org/10.1002/art.1780380107
21. van Gestel AM, Anderson JJ, van Riel PL, Boers M, Haagsma CJ, Rich B., *et al.* ACR and EULAR improvement criteria have comparable validity in rheumatoid arthritis trials. American College of Rheumatology European League of Associations for Rheumatology. *Journal of Rheumatology* 1999;**26**:705-11.
22. Stevenson M, Archer R, Tosh J, Simpson E, Everson-Hock E, Stevens J, *et al.* Adalimumab, etanercept, infliximab, certolizumab pegol, golimumab, tocilizumab and abatacept for the treatment of rheumatoid arthritis not previously treated with disease-modifying antirheumatic drugs and after the failure of conventional disease-modifying antirheumatic drugs only: systematic review and economic evaluation. *Health Technology Assessment* In press.
23. Bruce B, Fries JF. The Health Assessment Questionnaire (HAQ). *Clin Exp Rheumatol* 2005;**23**:S14-S8.
24. National Institute for Health and Clinical Excellence (NICE). *Adalimumab, etanercept, infliximab, certolizumab pegol, golimumab, tocilizumab and abatacept for rheumatoid arthritis not previously treated with DMARDs or after conventional DMARDs only have failed*; 2016.
25. National Institute for Health and Clinical Excellence (NICE). *Certolizumab pegol for the treatment of rheumatoid arthritis*; 2010.
26. National Institute for Health and Clinical Excellence (NICE). *Tocilizumab for the treatment of rheumatoid arthritis.*; 2012.
27. National Institute for Health and Care Excellence. *Certolizumab pegol for treating rheumatoid arthritis after inadequate response to a TNF inhibitor – final scope*. London: NICE; 2016.
28. UCB. *Certolizumab Pegol for treating rheumatoid arthritis after inadequate response to a TNF inhibitor — UCB Response to ERG Clarification Questions*. Berkshire; 2016.
29. CIMZIA. Annex I. Summary of Product Characteristics.
30. British National Formulary. British National Formulary (BNF) 71. 2016.
31. UCB. *Certolizumab Pegol for treating rheumatoid arthritis after inadequate response to a TNF inhibitor*. Berkshire; 2016.
32. Yamamoto K, Takeuchi T, Yamanaka H, Ishiguro N, Tanaka Y, Eguchi K, *et al.* Efficacy and safety of certolizumab pegol plus methotrexate in Japanese rheumatoid arthritis patients with an inadequate response to methotrexate: the J-RAPID randomized, placebo-controlled trial. *Modern rheumatology / the Japan Rheumatism Association* 2014;**24**:715-24. http://dx.doi.org/10.3109/14397595.2013.864224
33. Yamamoto K, Takeuchi T, Yamanaka H, Ishiguro N, Tanaka Y, Eguchi K, *et al.* Efficacy and safety of certolizumab pegol without methotrexate co-administration in Japanese patients with active rheumatoid arthritis: the HIKARI randomized, placebo-controlled trial. *Modern rheumatology / the Japan Rheumatism Association* 2014;**24**:552-60. http://dx.doi.org/10.3109/14397595.2013.843764
34. Weinblatt ME, Fleischmann R, Huizinga TWJ, Emery P, Pope J, Massarotti EM, *et al.* Efficacy and safety of certolizumab pegol in a broad population of patients with active

rheumatoid arthritis: results from the REALISTIC phase IIIb study. *Rheumatology* 2012;**51**:2204-14. http://dx.doi.org/10.1093/rheumatology/kes150

35. Furst DE, Shaikh SA, Greenwald M, Bennett B, Davies O, Luijtens K, *et al.* Two dosing regimens of certolizumab pegol in patients with active rheumatoid arthritis. *Arthritis Care Res (Hoboken)* 2015;**67**:151-60. http://dx.doi.org/10.1002/acr.22496
36. Curtis JR, Churchill M, Kivitz A, Samad A, Gauer L, Gervitz L, *et al.* A Randomized Trial Comparing Disease Activity Measures for the Assessment and Prediction of Response in Rheumatoid Arthritis Patients Initiating Certolizumab Pegol. *Arthritis & Rheumatology* 2015;**67**:3104-12. http://dx.doi.org/10.1002/art.39322
37. Schiff MH, von Kempis J, Goldblum R, Tesser JR, Mueller RB. Rheumatoid arthritis secondary non-responders to TNF can attain an efficacious and safe response by switching to certolizumab pegol: a phase IV, randomised, multicentre, double-blind, 12-week study, followed by a 12-week open-label phase. *Ann Rheum Dis* 2014;**73**:2174-7. http://dx.doi.org/10.1136/annrheumdis-2014-205325
38. European Medicines Agency Evaluation of Medicines for Human Use. *Assessment report for Cimzia.* ; 2009.
39. Kang Y, Park W, Park Y, Choe J, Bae S, Cho C, *et al.* Efficacy and safety of certolizumab pegol with concomitant methotrexate in Korean rheumatoid arthritis (RA) patients with an inadequate response to MTX. *Annals of the Rheumatic Diseases* 2012;**71**:666. http://dx.doi.org/10.1136/annrheumdis-2012-eular.501
40. Emery P, Keystone E, Tony HP, Cantagrel A, van Vollenhoven R, Sanchez A, *et al.* IL-6 receptor inhibition with tocilizumab improves treatment outcomes in patients with rheumatoid arthritis refractory to anti-tumour necrosis factor biologicals: results from a 24-week multicentre randomised placebo-controlled trial. *Annals of the Rheumatic Diseases* 2008;**67**:1516-23. http://dx.doi.org/10.1136/ard.2008.092932
41. Genovese MC, Fleischmann R, Furst D, Janssen N, Carter J, Dasgupta B, *et al.* Efficacy and safety of olokizumab in patients with rheumatoid arthritis with an inadequate response to TNF inhibitor therapy: outcomes of a randomised Phase IIb study. *Ann Rheum Dis* 2014;**73**:1607-15. http://dx.doi.org/10.1136/annrheumdis-2013-204760
42. Cohen SB, Emery P, Greenwald MW, Dougados M, Furie RA, Genovese MC, *et al.* Rituximab for rheumatoid arthritis refractory to anti-tumor necrosis factor therapy: Results of a multicenter, randomized, double-blind, placebo-controlled, phase III trial evaluating primary efficacy and safety at twenty-four weeks. *Arthritis Rheum* 2006;**54**:2793-806. http://dx.doi.org/10.1002/art.22025
43. Combe B, Dougados M, Logeart I, Lukas C, Mariette X, Puechal X, *et al.* AB0516 An open-label randomized controlled study to evaluate the efficacy of etanercept (ETN) versus rituximab (RTX), in patients with active rheumatoid arthritis previously treated with RTX and TNF blockers. *Annals of the Rheumatic Diseases* 2013;**71**:667. http://dx.doi.org/10.1136/annrheumdis-2012-eular.516
44. Genovese MC, Becker JC, Schiff M, Luggen M, Sherrer Y, Kremer J, *et al.* Abatacept for rheumatoid arthritis refractory to tumor necrosis factor alpha inhibition. *The New England journal of medicine* 2005;**353**:1114-23. http://dx.doi.org/10.1056/NEJMoa050524
45. Smolen JS, Kay J, Doyle MK, Landewe R, Matteson EL, Wollenhaupt J, *et al.* Golimumab in patients with active rheumatoid arthritis after treatment with tumour necrosis factor alpha inhibitors (GO-AFTER study): a multicentre, randomised, double-blind, placebo-controlled, phase III trial. *Lancet (London, England)* 2009;**374**:210-21. http://dx.doi.org/10.1016/s0140-6736(09)60506-7
46. Furst DE, Gaylis N, Bray V, Olech E, Yocum D, Ritter J, *et al.* Open-label, pilot protocol of patients with rheumatoid arthritis who switch to infliximab after an incomplete response to etanercept: the opposite study. *Ann Rheum Dis* 2007;**66**:893-9. http://dx.doi.org/10.1136/ard.2006.068304
47. Mease PJ, Cohen S, Gaylis NB, Chubick A, Kaell AT, Greenwald M, *et al.* Efficacy and safety of retreatment in patients with rheumatoid arthritis with previous inadequate response to tumor necrosis factor inhibitors: results from the SUNRISE trial. *J Rheumatol* 2010;**37**:917-27. http://dx.doi.org/10.3899/jrheum.090442

48. Gottenberg J, Brocq O, Perdriger A, Lassoued S, Berthelot J-M, Wendling D, *et al.* In the multicenter randomized controlled rotation or change trial, a non-TNF targeted therapy has a higher efficacy than a second anti-TNF at 3, 6 and 12 months. Arthritis & Rheumatology, abstract no. 108.
49. Chatzidionysiou K, Kristensen LE, Eriksson J, Askling J, van Vollenhoven R. Effectiveness and survival-on-drug of certolizumab pegol in rheumatoid arthritis in clinical practice: results from the national Swedish register. *Scandinavian journal of rheumatology* 2015;**44**:431-7. http://dx.doi.org/10.3109/03009742.2015.1026840
50. Bykerk VP, Cush J, Winthrop K, Calabrese L, Lortholary O, De Longueville M, *et al.* Update on the safety profile of certolizumab pegol in rheumatoid arthritis: an integrated analysis from clinical trials. *Annals of the Rheumatic Diseases* 2015;**74**:96-103. http://dx.doi.org/10.1136/annrheumdis-2013-203660
51. Yun H, Xie F, Delzell E, Levitan EB, Chen L, Lewis JD, *et al.* Comparative Risk of Hospitalized Infection Associated With Biologic Agents in Rheumatoid Arthritis Patients Enrolled in Medicare. *Arthritis Rheumatol* 2016;**68**:56-66. http://dx.doi.org/10.1002/art.39399
52. Simard JF, Arkema EV, Sundstrom A, Geborek P, Saxne T, Baecklund E, *et al.* Ten years with biologics: to whom do data on effectiveness and safety apply? *Rheumatology (Oxford, England)* 2011;**50**:204-13. http://dx.doi.org/10.1093/rheumatology/keq326
53. Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, *et al.* Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Controlled clinical trials* 1996;**17**:1-12.
54. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *Journal of epidemiology and community health* 1998;**52**:377-84.
55. Centre for Reviews Dissemination. *Systematic review: CRD's guidance for undertaking reviews in health care. CRD, University of York; 2009. The Cochrane Collaboration. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]*; 2011.
56. Higgins J, Thompson S, Spiegelhalter D. A re-evaluation of random-effects meta-analysis. *J R Statistis Soc A* 2009;**172**:137-59.
57. Curtis JR, Churchill M, Kivitz A, Samad A, Gauer L, Gervitz L, *et al.* A Randomized Trial Comparing Disease Activity Measures to Assess and Predict Response in Rheumatoid Arthritis Patients Initiating Certolizumab Pegol. *Arthritis Rheumatol* 2015; 10.1002/art.39322. http://dx.doi.org/10.1002/art.39322
58. Higgins J. Commentary: Heterogeneity in meta-analysis should be expected and appropriately quantified. . *International Journal of Epidemiology* 2008;**37**:1158-60.
59. Scottish Intercollegiate Guidelines Network. *Search Filters*. Healthcare Improvement Scotland; 2015. URL: http://www.sign.ac.uk/methodology/filters.html (Accessed 29/04/16, 2016).
60. Pennington B, Davis S. Mapping from the Health Assessment Questionnaire to the EQ-5D: The Impact of Different Algorithms on Cost-Effectiveness Results. *Value in Health* 2014;**17**:762-71. http://dx.doi.org/http://dx.doi.org/10.1016/j.jval.2014.11.002
61. Dakin H. Review of studies mapping from quality of life or clinical measures to EQ-5D: an online database. *Health and Quality of Life Outcomes* 2013;**11**:1-6. http://dx.doi.org/10.1186/1477-7525-11-151
62. National Institute for Health and Care Excellence. *Guide to the methods of technology appraisal*; 2013.
63. Keystone E, Landewé R, van Vollenhoven R, Combe B, Strand V, Mease P, *et al.* Long-term safety and efficacy of certolizumab pegol in combination with methotrexate in the treatment of rheumatoid arthritis: 5-year results from the RAPID 1 trial and open-label extension. *Annals of the Rheumatic Diseases* 2014;**73**:2094-100. http://dx.doi.org/10.1136/annrheumdis-2013-203695
64. Smolen J, Landewé RB, Mease P, Brzezicki J, Mason D, Luijtens K, *et al.* Efficacy and safety of certolizumab pegol plus methotrexate in active rheumatoid arthritis: the RAPID 2 study. A

randomised controlled trial. *Annals of the Rheumatic Diseases* 2009;**68**:797-804. http://dx.doi.org/10.1136/ard.2008.101659

65. Hernández Alava M, Wailoo A, Wolfe F, Michaud K. The relationship between EQ-5D, HAQ and pain in patients with rheumatoid arthritis. *Rheumatology* 2013;**52**:944-50. http://dx.doi.org/10.1093/rheumatology/kes400
66. Edwards CJ, Arden NK, Fisher D, Saperia JC, Reading I, Van Staa TP, *et al.* The changing use of disease-modifying anti-rheumatic drugs in individuals with rheumatoid arthritis from the United Kingdom General Practice Research Database. *Rheumatology* 2005;**44**:1394-8. http://dx.doi.org/10.1093/rheumatology/kei024
67. Office for National Statistics. *Population Estimates for UK, England and Wales, Scotland and Northern Ireland, Mid-2014.*; 2014.
68. Norton S, Sacker A, Dixey J, Done J, Williams P, Young A. Trajectories of functional limitation in early rheumatoid arthritis and their association with mortality. *Rheumatology* 2013;**52**:2016-24. http://dx.doi.org/10.1093/rheumatology/ket253
69. Michaud K, Vera-Llonch M, Oster G. Mortality Risk by Functional Status and Health-related Quality of Life in Patients with Rheumatoid Arthritis. *J Rheumatol* 2012;**39**:54-9. http://dx.doi.org/10.3899/jrheum.110491
70. Brennan A, Bansback N, Nixon R, Madan J, Harrison M, Watson K, *et al.* Modelling the cost effectiveness of TNF-α antagonists in the management of rheumatoid arthritis: results from the British Society for Rheumatology Biologics Registry. *Rheumatology* 2007;**46**:1345-54. http://dx.doi.org/10.1093/rheumatology/kem115
71. Bansback N, Young A, Brennan A, Dixey J. A prognostic model for functional outcome in early rheumatoid arthritis. *J Rheumatol* 2006;**33**:1503-10. http://www.jrheum.org/content/33/8/1503.abstract
72. British National Formulary. British National Formulary (BNF) 64. 2012.
73. Chakravarty K, McDonald H, Pullar T, Taggart A, Chalmers R, Oliver S, *et al.* BSR/BHPR guideline for disease-modifying anti-rheumatic drug (DMARD) therapy in consultation with the British Association of Dermatologists. *Rheumatology* 2008;**47**:924-5. http://dx.doi.org/10.1093/rheumatology/kel216a
74. Deparment of Health. *NHS reference Costs 2014 to 2015*. 2015. URL: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/480791/2014-15_National_Schedules.xlsx (Accessed 24 March, 2016).
75. Personal Social Services Research Unit. Unit Costs of Health and Social Care. 2015, http://www.pssru.ac.uk/project-pages/unit-costs/.
76. Wiles NJ, Symmons D. *Resource Use Within the Norfolk Arthritis Register (NOAR) Cohort During the First Five Years of Disease. Report for Roche.* London: NICE; 2005.
77. Department of Health. *National Schedule of Reference Costs Year: 2010-11 - NHS Trusts and PCTs Combined HRG Data.*; 2013.
78. Office for National Statistics. *Consumer Price Index.*; 2016. URL: https://www.ons.gov.uk/search?q=consumer+price+index (Accessed 4 May, 2016).
79. Kobelt G, Jönsson L, Lindgren P, Young A, Eberhardt K. Modeling the progression of rheumatoid arthritis: A two-country model to estimate costs and consequences of rheumatoid arthritis. *Arthritis & Rheumatism* 2002;**46**:2310-9. http://dx.doi.org/10.1002/art.10471
80. Drummond M, Sculpher M, Torrance G, O'Brien B, G S. *Methods for the economic evaluation of health care programmes.* 3rd edn. Oxford: Oxford University Press; 2005.
81. Weinstein MC, O'Brien B, Hornberger J, Jackson J, Johannesson M, McCabe C, *et al.* Principles of Good Practice for Decision Analytic Modeling in Health-Care Evaluation: Report of the ISPOR Task Force on Good Research Practices — Modeling Studies. *Value in Health* 2014;**6**:9-17. http://dx.doi.org/10.1046/j.1524-4733.2003.00234.x
82. National Institute for Health and Care Excellence. *Adalimumab for treating moderate to severe hidradenitis suppurativa – final scope*. London: NICE; 2015.
83. Ades A, Higgins J. The interpretation of random-effects meta-analysis in decision models. *Medical Decision Making* 2005;**25**:646-54.
84. Bansback N, Marra C, Tsuchiya A, Anis A, Guh D, Hammond T, *et al.* Using the health assessment questionnaire to estimate preference-based single indices in patients with

rheumatoid arthritis. *Arthritis Care & Research* 2007;**57**:963-71. http://dx.doi.org/10.1002/art.22885

85. European Medicines Agency. *Summary of product characteristics – tocilizumab (RoActemra)*. London; 2015.
86. *MIMS Online Source*. 2016. URL: http://www.mims.co.uk/ (Accessed 3 May, 2016).
87. Burmester GR, Rubbert-Roth A, Cantagrel A, Hall S, Leszczynski P, Feldman D, *et al.* A randomised, double-blind, parallel-group study of the safety and efficacy of subcutaneous tocilizumab versus intravenous tocilizumab in combination with traditional disease-modifying antirheumatic drugs in patients with moderate to severe rheumatoid arthritis (SUMMACTA study). *Annals of the Rheumatic Diseases* 2014;**73**:69-74. http://dx.doi.org/10.1136/annrheumdis-2013-203523

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

Single Technology Appraisal (STA)

Certolizumab pegol for treating rheumatoid arthritis after inadequate response to a TNF inhibitor: A Single Technology Appraisal [ID824]

May 2016

UCB Response to ERG Report

UCB welcomes the opportunity to review the ERG's report on certolizumab pegol (CZP) for treating rheumatoid arthritis after inadequate response to a TNF inhibitor. UCB welcomes the conclusions of the ERG, similar to those from the UCB submission, that CZP is a cost-effective option as combination therapy with methotrexate (MTX) or monotherapy, in populations B and C of the submission, where currently other anti-TNFs are recommended by NICE.

A summary of the key points of inaccuracy identified in the ERG report reference are provided below and detailed thereafter.

1. *Dealing with uncertainty:*
   UCB believe there to be a number of inaccuracies described on how the uncertainty was explored within the original UCB submission, through conducting a range of scenario analyses, and evaluated all possible evidence identified from a robust systematic review of the literature.

2. *Conclusion on biosimilars*
   UCB strongly objects to the inclusion of statements on the positioning of other interventions, which are not in scope of the current appraisal. As per the NICE guidance on Single Technology Appraisal processes, the scope of this appraisal is the evaluation of the clinical and cost effectiveness of CZP within the context of the final appraisal remit and objectives, and not whether biosimilar etanercept (or other products) should be preferentially positioned to CZP.

Please find below the list of the factual inaccuracies identified in the ERG report and the justifications for amendments proposed.

Key Issues

Issue 1 Dealing with Uncertainty

Description of problem	Description of proposed amendment	Justification for amendment	ERG's response
Point 1: ERG report, Executive Summary, Section 1.5 (page 17): "The ERG believes that the assumptions made by the company to overcome the lack of evidence did not evaluate fully the uncertainty and that therefore the outcomes of the model should be interpreted with caution."	We suggest that the ERG refers to the UCB submitted scenario analysis presented in the original Company Submission (CS) and revises this sentence.	UCB believes that the quoted sentence in the Executive Summary is too strong and does not accurately reflect the analyses submitted by UCB. The uncertainty around the lack of published evidence for comparators was explored in the UCB original submission, through an extreme scenario analyses where a class effect of anti-TNFs was investigated in all three populations assessed (CS page 249).	Amended sentence to: "The ERG believes that the assumptions made by the company to overcome the lack of evidence for its base case analysis did not evaluate fully the uncertainty and that therefore the results of the base case analysis should be interpreted with caution"
Point 2: ERG report, Section 1.6.2 (page 18) states: "The absence of evidence on the efficacy of bDMARDs in patients who have had an inadequate response to two or more bDMARDs introduced considerable uncertainty in the model. The company adopted simplifying assumptions in order to model the efficacy of bDMARDs in subsequent treatments, the impact of which is unknown, although results comparing elongated and standard sequences appeared	We kindly ask that the ERG clarifies in Section 1.6.2 that the lack of evidence presented for competitors is due to the lack of published evidence and thus not a limitation of the analysis conducted by UCB.	We think that the statement is currently misleading and needs to be amended to accurately reflect the UCB submission. As lack of evidence for competitors such as TOC have been previously acknowledged and accepted by NICE, eg. in NICE TA247[1], we kindly request that the ERG clarifies that the lack of evidence presented for competitors is due to the lack of published evidence and thus not a limitation of the analysis conducted by UCB.	Amended the sentence to: "The lack of published evidence on the efficacy of competitor bDMARDs in patients who have had an inadequate response to two or more bDMARDs introduced considerable uncertainty in the model. The company adopted simplifying assumptions in order to model the efficacy of bDMARDs in subsequent treatments, the impact of which is unknown, although results comparing elongated and standard sequences appeared implausible. As such, the ERG

[1] NICE: https://www.nice.org.uk/guidance/ta247

implausible. As such, the ERG believes that the credibility of comparisons of sequences of different lengths within the model is limited."			believes that the credibility of comparisons of sequences of different lengths within the model is limited."
Point 3: Accordingly, we believe that the areas of uncertainty raised in Section 7 (points 1. and 2. on page 164) are too strongly worded.	We request revision to acknowledge similar data having been accepted eg. in TA195: TA195, Section 4.3.13, page 34: "The Committee considered that there are significant limitations in the evidence base available for this appraisal and that the relative clinical effectiveness of TNF inhibitors after the failure of a first TNF inhibitor remains uncertain."	We kindly ask for ERG to consider rewording these statements to reflect that these are limitations due to lack or limited published evidence of the efficacy data of comparators, that have already been highlighted as limitations in the previous NICE TA195, and are not specific to the UCB submission. Furthermore these limitations were already acknowledged during the scoping meeting of the current appraisal by the AG.	Added following sentence in Section 7: "This limitation had already been highlighted by the AC of TA195.[25]"

Issue 2 Conclusion on biosimilars

Description of problem	Description of proposed amendment	Justification for amendment	ERG's response
ERG report Section 7 (paragraph 3 on page 164): "As these will remain options for treatment in Populations B and C a positive recommendation for CZP + MTX would allow an alternative if ETA biosimilar was not appropriate."	Based on the justification provided, UCB would strongly request the removal of this text from the ERG document.	UCB strongly objects to the inclusion of the statement made on page 167. As a Single Technology Appraisal and based on the final appraisal remit / objective ("*To appraise the clinical and cost effectiveness of certolizumab pegol within its marketing authorisation for treating rheumatoid arthritis after inadequate response to a TNF inhibitor*.") UCB believe this appraisal should only permit consideration of the clinical and cost effectiveness of intervention under scope, that is CZP, within the context of	Amended sentence to: "The latter two interventions will remain options recommended by NICE for treatment in Populations B and C."

		the abovementioned remit, and not whether biosimilar etanercept (or other products) should be preferentially positioned to CZP as the passage effectively implies. UCB also believes this text is contrary to the guidance provided by NICE's biosimilars position statement, which states that biosimilar products will usually be considered as interventions within the context of an MTA in parallel with their reference product in the indication under consideration, something not applicable to this Single Technology Appraisal. Based on the justification provided, UCB would strongly request the removal of this text from the ERG document.	

Further Issues

Issue 3 Clinical-effectiveness: Description of SWITCH Study

Description of problem	Description of proposed amendment	Justification for amendment	ERG's response
Point 1: In the executive summary (Section 1.2, page 13), the following sentence does not indicate that in the SWITCH trial, both arms received CZP after week 12: "Whilst CZP-treated patients in the SWITCH trial had a more favourable EULAR response at	We request for the sentence to be changed to: "CZP-treated patients in the SWITCH trial had a more favourable EULAR response at week 12 compared to placebo, which was the end of the double-blind period. Results at week 24, end of the OLE period,	The statement is inaccurate with respect to the SWITCH study design. It is inappropriate to infer that the CZP and placebo have similar efficacy at week 24, when all patients randomized to either CZP or PBO, receive CZP after week 12, until week 24. The text should clearly reflect that during the OLE phase, from week 12 to week 24 all	Amended the sentence to: "Whilst CZP-treated patients in the SWITCH trial had a more favourable EULAR response at week 12 (end of the double-blind period), results at week 24 (end of the OLE period, during which all patients received CZP) were broadly comparable between study

week 12, results at week 24 were broadly comparable between CZP and PBO treatment groups."	during which all patients received CZP, were similar. "	patients received CZP after week 12.	groups."
Point 2: Similarly, in Section 4.2.5.2 (page 64), the ERG omits this information by stating: "Interestingly, analyses reported on ClinicalTrials.gov indicated that the proportions of ACR20 and ACR50 responders were actually lower in CZP-treated patients (54.5% and 27.3%) compared with PBO subjects (62.5% and 37.5%) at week 24 […]." Further, the patient numbers in the OLE phase were not reported here.	We request an amendment of the statement as follows, to clarify CZP treatment in the OLE phase as well as to include the patient numbers: "At week 24, end of the OLE period, during which all patients received CZP regardless of the initial randomization, proportions of ACR20 and ACR50 responders were 62.5% (5/8 patients) and 37.5% (3/8 patient) versus 54.5% (12/22 patients) and 27.3% (6/22 patients), in the respective patient groups treated with CZP during the OLE phase."	The statement is inaccurate with respect to the SWITCH study design. The statement by the ERG omits that patients in the "PBO" arm were receiving CZP after Week 12, and thus were receiving CZP at week 24, thus statements of efficacy of CZP versus PBO can not be made at Week 24, since all patients were treated with CZP at week 24.	Amended sentence to: "Interestingly, analyses reported on ClinicalTrials.gov indicated that the proportions of ACR20 and ACR50 responders at week 24 (end of the 12-week OLE period, during which all patients received CZP) were lower in patients initially treated with CZP (54.5% and 27.3%) compared with patients initially treated with PBO (62.5% and 37.5%)"

Issue 4 Cost-effectiveness: Population C NMA Results

Description of problem	Description of proposed amendment	Justification for amendment	ERG's response
Point 1: The following sentence in Section 1.2 (page 14) is incorrect: "The results from the Bayesian network meta-analysis (NMA) within the CS indicated that … for population C [redacted]	We request to remove this statement regarding population C. Instead it could be stated that: "No NMA could be performed for population C due to lack of data for the comparators."	The statement is inaccurate, since no NMA was performed for population C due to lack of published evidence for comparators (CS Section 4.10.5).	Amended sentence to: "The results from the Bayesian network meta-analysis (NMA) within the CS indicated that: for population A [redacted]; and for population B [redacted]

[redacted]			[redacted]. No NMA could be performed for population C due to lack of data for the comparators."
Point 2: Further, a statement about "+MTX" in Section 1.2 (page 14) and Section 7 (page 164) is incorrect: "For Population C, the probabilistic ICER of CZP + MTX versus ETA(bio) monotherapy […]"	We request for the statement of "+MTX" for population C to be removed.	Population C is monotherapy and therefore any statements where '+ MTX' has been included are incorrect.	Amended sentence to: "For Population C, the probabilistic ICER of CZP monotherapy versus ETA(bio) monotherapy […]"

Issue 5 Cost-effectiveness: Duration of Treatment

Description of problem	Description of proposed amendment	Justification for amendment	ERG's response
Point 1: The following sentence in Section 1.5 (page 17) does not accurately reflect the data on treatment duration in the literature: "The ERG notes that the company assumed the same treatment duration for all bDMARDs despite evidence suggesting different treatment durations for different bDMARDs."	We request amendment of the paragraph with added wording underlined, to read: "The ERG notes that the company assumed the same treatment duration for all bDMARDs, <u>in their base case</u> despite evidence suggesting different treatment durations for different bDMARDs. <u>However, both scenarios of shorter and longer treatment duration were addressed in the CS and tested with scenario analysis.</u>"	The statement does not accurately reflect the submitted evidence and the assumptions made in the cost-effectiveness model. We acknowledge that there is evidence to suggest different treatment durations for different bDMARDs, however there is no consensus in the literature on the duration, with evidence suggesting both longer and shorter treatment durations for anti-TNFs compared to biologic non anti-TNFs. Given the great uncertainty over treatment duration difference, a conservative assumption was made in the base case, where all biologics were assumed to have a similar treatment duration and a scenario analysis was further conducted where both longer and shorter durations were tested to address	Amended sentence to: "The ERG notes that the company assumed the same treatment duration for all bDMARDs for its base case analysis despite evidence suggesting different treatment durations for different bDMARDs."

		this uncertainty in the most appropriate way (CS Section 5.8.3).	
Point 2: In section 5.2.9 (page 132) of the ERG report it is stated that: "The source of this value was not provided in the CS."	We request the removal of this sentence.	The source of this value (Ramiro 2015) is provided on page 199 of the CS.	Sentence removed from the report.

Issue 6 Cost-effectiveness: Treatment Sequence Clarification

Description of problem	Description of proposed amendment	Justification for amendment	ERG's response
The following statement in Section 1.7 (page 20) implies that specific statements were made in the CS which were not: "[redacted]" The same statement is repeated in Section 7 (page 165).	We request for it to be made clear in both instances that the second part of the sentence ("[redacted]") is an interpretation made by the ERG and was not stated in the CS. We suggest the following revision: "The ERG interpret data from [redacted]	The viewpoints of the ERG should be clearly delineated from statements that come from the CS.	Amended sentence to: "[redacted]"

Issue 7 Background: Use and Description of TA375

Description of problem	Description of proposed amendment	Justification for amendment	ERG's response
Point 1: In Section 2.2 (page 24), the ERG states: "It is noted that the recommendations from TA375 have not been incorporated into the CS."	We kindly request the removal of this sentence.	TA375 solely provides recommendations on first line anti-TNF use, while the scope of this STA is to "appraise the clinical and cost effectiveness of certolizumab pegol […] after inadequate response to a TNF inhibitor". The CS refers to guidances and pathways in Sections 1.1 and 3.3.	No change made. Section 2.2 refers to current service provision for the whole pathway of RA. The ERG still believes the recommendations from TA375 were not incorporated in some parts of the CS, e.g. in Figure 2.
Point 2: In Section 2.2 (page 25), the following sentence is incorrect: "*TA195, which for all bDMARDs excluding RTX was updated in* TA375, states that bDMARD treatment *after the failure of a TNFi* should be continued only if there is an adequate response (defined as an improvement in the DAS28 score of at least 1.2 points) at initiation of treatment and as long as this adequate response is maintained."	We request for the section in italics to be removed to make the sentence accurate. Therefore, the sentence should read: "TA375 states that bDMARD treatment should be continued only if there is an adequate response (defined as an improvement in the DAS28 score of at least 1.2 points) at initiation of treatment and as long as this adequate response is maintained."	TA375 provides recommendations solely on first-line anti-TNF use, and it was not within its scope to update TA195, as the ERG text implies.	Amended sentence to: "TA195 states that bDMARD treatment after the failure of a TNFi should be continued only if there is an adequate response (defined as an improvement in the DAS28 score of at least 1.2 points) at initiation of treatment and as long as this adequate response is maintained."

Issue 8 Decision Problem: Cost of CZP

Description of problem	Description of proposed amendment	Justification for amendment	ERG's response
In Section 3.2 (page 28)	As per the company submission	Clarifies the correct cost of CZP.	Amended sentence to:

the ERG incorrectly reports the cost of CZP as **£375.50** per 200 mg pre-filled syringe.	(page 31), the cost of CZP is **£357.50** per 200 mg pre-filled syringe. We request the ERG to correct their report accordingly.		"The price of a 200mg syringe prefilled with CZP is £357.50."

Issue 9 Clinical-effectiveness: Patient Disposition in J-RAPID and HIKARI

Description of problem	Description of proposed amendment	Justification for amendment	ERG's response
In Section 4.2.3 (page 50), the ERG gives a misleading statement regarding completion rates, saying that: "[…] only **35.5%** and 15.8% of PBO group patients completed (with most withdrawing *at week 16 due to lack of efficacy*)."	We ask for the section in italics to be replaced to clarify that withdrawal at Week 16 was scheduled as per trial design. Furthermore, the completion rate for the PBO group of J-RAPID should be corrected to **32.5%** (25/77) as per CS Figure 14. Therefore, the sentence should read: "[…] only **32.5%** and 15.8% of PBO group patients completed (with most exiting via scheduled withdrawal due at Week 16, due to not achieving an ACR 20 at weeks 12 and 14, as per study design)."	Clarifies that withdrawal was scheduled as per study design and the criteria. Provides accurate rate of completion for PBO arm in J-RAPID.	Sentence amended to: "only 32.5% and 15.8% of PBO group patients completed (with most withdrawing at week 16 due to not achieving an ACR 20 at weeks 12 and 14, as per study design)."

Issue 10 Clinical-effectiveness: REALISTIC: Patient Numbers in TNFi-experienced Subgroups

Description of problem	Description of proposed amendment	Justification for amendment	ERG's response
In Section 4.2.5.1, the ERG incorrectly states that the reason for TNFi-experienced patient subgroups in CS tables for	We kindly as for the removal of these statements in Section 4.2.5.1 (page 58) as well as in the legend of Table	The CS Section 4.3.1 (page 55) clearly states that next to the subgroups on **a)** monotherapy and **b)** combination with MTX with or without further concomitant	Sentences deleted from report

REALISTIC not tallying to the total patient numbers is unclear. page 58: "[…] did not appear to tally with the total number of TNFi-experienced patients for reasons that are unclear to the ERG (ie. [redacted] Table 10 (page 61): "[…] ACR analysis subgroups did not tally with the total number of TNFi-experienced patients for reasons unclear."	10 (page 61).	cDMARDs, a third subgroup of patients (**c)** CZP in combination with concomitant cDMARDs other than MTX) was included in the study. The latter subgroup is out of the scope of this STA and efficacy data was thus not presented for this subgroup, resulting in the patient numbers presented. Patient numbers in each group were [redacted], tallying up to 320 patients on CZP and 80 on PBO.	

Issue 11 Clinical-effectiveness: Description of DOSEFLEX Study

Description of problem	**Description of proposed amendment**	**Justification for amendment**	**ERG's response**
Point 1: In Section 4.2.6 (page 71) the following sentence does not indicate that in the DOSEFLEX trial, all patient arms received CZP in the run-in phase: "In DOSEFLEX[redacted]TNFi-experienced patients *in the CZP + MTX and PBO + MTX treatment arms* achieved EULAR good or moderate response rates at week 16 (the end of open-label run-in)."	We ask for the text in italics to be removed. The sentence should thus read: "In DOSEFLEX, TNFi-experienced patients treated with CZP achieved EULAR good or moderate response rates at week 16 (the end of open-label run-in)."	The current text does not accurately reflect the DOSEFLEX study design. As per study design, the run-in phase of DOSEFLEX comprised of a single treatment arm, where all patients were treated with CZP. It is inappropriate to infer that patients were treated with PBO during the run-in phase. It should thus be clearly stated that all patients received CZP during the run-in phase (Weeks 0–16) and were randomised to CZP or PBO thereafter, resulting in the observed treatment effects.	Sentence amended to: "In DOSEFLEX[redacted]proportions of TNFi-experienced patients in the study arms achieved EULAR good or moderate response rates at week 16 (the end of open-label run-in where both groups were treated with CZP + MTX)"
Point 2: The ERG makes inaccurate	Text in italics to be replaced by "all CZP-treated patients during the		Sentence in page 73 amended to:

statements on: Page 73: "*All treatment groups* in the DOSEFLEX tria[redacted]HAQ-DI during the 16 week open label run-in period (see Figure 6)." Page 78: "Fatigue [redacted]to week 16 across *all treatment groups during the open-label CZP run-in* (Figure 10)."	single arm open-label run-in phase". Change sentence to: Page 73: "*CZP treated patients* in the DOSEFLEX trial showed [redacted]during the 16 week open label run-in period (see Figure 6)." Page 78: "Fatigue[redacted]week 16 with CZP *during the open-label run-in* [redacted]"		"Both study groups in the DOSEFLEX trial [redacted] HAQ-DI during the 16 week open label CZP run-in period (see Figure 6)." Sentence in page 78 amended to: "Fatigue[redacted]to week 16 with CZP during the open-label run-in in all study groups (Figure 10)"
Point 3: In Section 4.2.9 (page 77), there is an inaccurate statement about the randomised phase: "However, during the randomised controlled period (to week 34), […]"	Change sentence to: Page 77: "However, during the randomised controlled period (**week 18** to week 34), […]"	Clarifies statement regarding randomised phase.	Sentence amended to: "However, during the randomised controlled period (week 18 to week 34)"

Issue 12 Clinical-effectiveness: REALISTIC: Physical Function

Description of problem	Description of proposed amendment	Justification for amendment	ERG's response
In Section 4.2.7 (page 72), the ERG incorrectly refers to "mean HAQ-DI **scores**" instead of mean **change** in the paragraph below: "In the TNFi-experienced patients in REALISTIC, the mean HAQ-DI **scores** in both CZP[redacted]and PBO [redacted]treatment groups showed reductions from baseline to week 12 [redacted]The CS states that for HAQ-DI	Replace "mean score" with "mean change" in the final paragraph on page 72: "In the TNFi-experienced patients in REALISTIC, the mean HAQ-DI **change** in both CZP [redacted]and PBO [redacted]treatment groups showed[redacted]from baseline to week 12 (see Figure 5). The CS states that for HAQ-DI[redacted]	Clarifies the type of scores provided.	Sentence amended to: In the TNFi-experienced patients in REALISTIC, the mean HAQ-DI **changes** in both CZP [redacted] and PBO [redacted]treatment group showed reductions from baseline to week 12 [redacted]The CS states that for HAQ-DI[redacted]

score there was a [redacted]			

Issue 13 Clinical-effectiveness: Reporting of Pain Data

Description of problem	Description of proposed amendment	Justification for amendment	ERG's response
In Section 4.2.9 (page 76) the ERG incorrectly states that pain as an outcome was only covered briefly in the CS.	As per the CS (Sections 8.7.3.1 and 8.8.3.1, pages 25 and 29 of the CS appendices), pain, as evaluated using the patient's assessment of pain VAS, was reported from REALISTIC and DOSEFLEX. UCB would suggest that the ERG's interpretation of the amount of coverage given to pain evaluation is currently misleading and kindly requests that the ERG notes that pain outcomes were presented from the main CZP trials.	Clarifies the reporting of pain outcomes from CZP trials, to accurately reflect the UCB submission.	Sentence omitted from the report

Issue 14 Clinical-effectiveness: Safety Endpoints

Description of problem	Description of proposed amendment	Justification for amendment	ERG's response
Point 1: In Section 4.2.16 (page 87) the ERG has consolidated all of the safety endpoints from DOSEFLEX, REALISTIC and SWITCH. The outcomes are reported as "incidence rate per 100	We would propose adaptation of the heading of "AEs, incidence rate/100 PY (n, patient %)" to "AEs, incidence rate/100 PY (n, patient %) for DOSEFLEX and REALISTIC, and to "n (patient %)" for SWITCH for clarity of reporting of data.	Clarifies the correct safety endpoints across the REALISTIC and SWITCH studies, to accurately reflect the submitted evidence.	Set a "AEs, n (patient %)" header for the SWITCH trial data.

patient years (n, %)". This is correct reporting of information for DOSEFLEX and REALISTIC, but not for SWITCH, where the data reported is as n (%) only.			
Point 2: In Section 4.2.16.2 (page 85), the ERG report the proportion of patients treated with CZP reporting a serious infection in the REALISTIC study at **4.3%.**	As reported on page 149 of the CS, **2.6%** of patients receiving CZP up to Week 12 of the double-blind phase experienced a serious infection.		Sentence amended to: "Specific reported reasons for SAEs in the DOSEFLEX CZP groups included serious infections (CZP 200 mg Q2W 4.3%), cardiac disorders (CZP 200 mg Q2W, 1.4%), and musculoskeletal/connective tissue disorders (CZP 200 mg Q2W 2.9%, CZP 400 mg Q4W, 1.4%)."
Point 3: In Section 4.2.16.3 (page 85), the ERG report: "In REALISTIC, 10% of PBO group subjects had AEs leading to withdrawal, with no cases in the CZP 200 mg Q2W arm"	For consistency with the evidence on page 150 of UCB's submission document, we would suggest replacing the noted text with the following revision: "During the first 12 weeks of REALISTIC, 3.8% of PBO group subjects had AEs leading to withdrawal, with 4.7% of cases in the CZP 200 mg Q2W arm".		Sentence amended to: "During the first 12 weeks of REALISTIC, 3.8% of PBO group subjects had AEs leading to withdrawal, with 4.7% of cases in the CZP 200 mg Q2W arm"

Issue 15 Cost-effectiveness: Age-adjusted Utilities

Description of problem	Description of proposed amendment	Justification for amendment	ERG's response
In Section 5.3.2 (page 145) of the ERG report, the ERG make the following statement: "Due to the company's assumption that changes in HAQ score affects mortality, different treatments can result in different simulated mean survival. As such, the ERG believes that the utilities within the model should have been age-adjusted to account for reduced utility in the older population."	We kindly request for the statement to be revised.	UCB believes that the current statement does not accurately reflect the submitted evidence. Age-adjusted utilities have been indirectly included within the UCB model, as HAQ score increases with age in the model (for subsequent therapy cDMARDs and palliative care; Section 5.3.6, page 203 of the company submission).	No change made. The ERG still believes that utilities should have been adjusted. The company seems to imply that bDMARDs can stop the deleterious effects of ageing if it is assumed that the HAQ score remains constant whilst on treatment.

Issue 16 Cost-effectiveness: EQ-5D Data from REALISTIC

Description of problem	Description of proposed amendment	Justification for amendment	ERG's response
In Section 7 (page 148) of the ERG report it is stated: "The ERG considers that it would have been preferable to *use EQ-5D scores collected in REALISTIC* and that should there be significant differences between the two populations, the utilities currently used in the model would be biased."	UCB requests the deletion of the statement.	The statement made does not accurately reflect what was collected in REALISTIC, as indicated in the UCB submission. The statement implies that EQ-5D data from REALISTIC should have been used, however, EQ-5D was not collected in REALISTIC (Table 11, page 52 of the company	Added following sentence: "Unfortunately EQ-5D scores were not collected in REALISTIC."

		submission) and therefore this statement cannot be made.	

Issue 17 Cost-effectiveness: Typo on Page 154

Description of problem	Description of proposed amendment	Justification for amendment	ERG's response
In Section 5.4.1 (page 151) of the ERG report the incorrect sequence is given: "[…] that the currently recommended pathway (Sequence 4) dominates the same sequence if RTX is replaced with CZP (Sequence **4**)."	We request correction to: "[…] that the currently recommended pathway (Sequence 4) dominates the same sequence if RTX is replaced with CZP (Sequence **3**)."	Correct typo	Corrected

Issue 18 Cost-effectiveness: Incorrect Statement on Lines of Therapy

Description of problem	Description of proposed amendment	Justification for amendment	ERG's response
An incorrect statement is made by the ERG in Section 5.4.1 (page 151): "It is noticeable that Sequence 4, which includes only three lines of biologics, dominates Sequences 1 and 3, both of which include four lines of biologics." Sequence 3 replaces RTX with CZP and therefore only has 3 lines of therapy.	We request to remove reference to sequence 3 in this sentence. We suggest this could be changed to: "It is noticeable that Sequence 4, which includes only three lines of biologics, dominates Sequence 1, which includes four lines of biologics."	Sequence 3 replaces RTX with CZP and therefore only has 3 lines of therapy, making the sentence is incorrect. If the ERG wanted to make a comment about Sequence 4 (rather than sequence 3) having 4 lines of biologics, then the statement about dominance would need to change as CZP after RTX was not dominated by the original sequence (Sequence 1) with RTX alone.	"It is noticeable that Sequence 4, which includes only three lines of biologics, dominates Sequence 1, which includes four lines of biologics."

Certolizumab pegol for treating rheumatoid arthritis after inadequate response to a TNF inhibitor: A Single Technology Appraisal

Produced by	School of Health and Related Research (ScHARR), The University of Sheffield
Authors	Matt Stevenson, Professor of Health Technology Assessment, ScHARR, University of Sheffield, Sheffield, UK
	Inigo Bermejo, Research Associate, ScHARR, University of Sheffield, Sheffield, UK
	Rachel Archer, Research Fellow, ScHARR, University of Sheffield, Sheffield, UK
	John Stevens, Reader in Decision Science, ScHARR, University of Sheffield, Sheffield, UK
	Edward Goka, Research Associate, ScHARR, University of Sheffield, Sheffield, UK
	Mark Clowes, Information Specialist, ScHARR, University of Sheffield, Sheffield, UK
	David Scott, Department of Rheumatology, King's College Hospital NHS Foundation Trust, London, UK
	Adam Young, Department of Rheumatology, West Hertfordshire Hospital NHS Trust, Hertfordshire, UK
Correspondence Author	Matt Stevenson, Professor of Health Technology Assessment, ScHARR, University of Sheffield, Sheffield, UK
Date completed	Date completed (17/05/2016)

Source of funding: This report was commissioned by the NIHR HTA Programme as project number 15/06/07.

[redacted] at week 12 than those on PBO [redacted] The proportion of patients who achieved a good EULAR response was [redacted] for patients receiving CZP + MTX than CZP monotherapy [redacted] EULAR responses were also observed for CZP + MTX-treated patients in DOSEFLEX and J-RAPID versus PBO. Whilst CZP-treated patients in the SWITCH trial had a more favourable EULAR response at week 12 (end of the double-blind period), results at week 24 (end of the OLE period, during which all patients received CZP) were broadly comparable between CZP and PBO study groups. TNFi-experienced subjects in REALISTIC who received CZP as monotherapy experienced [redacted] EULAR good response compared with PBO at 12 weeks [redacted] proportion of those treated with CZP monotherapy had a EULAR moderate response [redacted] In HIKARI, more CZP monotherapy-treated TNFi-experienced patients achieved good [redacted] or moderate [redacted] EULAR responses at week 12 vs. PBO [redacted] CZP benefits were [redacted] for good EULAR response [redacted] but were more similar between groups for moderate EULAR response [redacted] EULAR data at 3 months were meta-analysed to compare the effects of CZP (both in combination with MTX and as monotherapy) with PBO. As previously for ACR response, the forest plots indicate [redacted] effects for the intervention groups in terms of EULAR response, although [redacted]) this effect appears [redacted] CZP as monotherapy and these results were [redacted]

Four RCTs (REALISTIC, DOSEFLEX, PREDICT and SWITCH) presented Health Assessment Questionnaire Disability Index (HAQ-DI) scores (or MD-HAQ/M-HAQ in the case of PREDICT). CZP-treated patients [redacted] No data for the outcomes of joint damage or radiological progression were included in the CS. Patients receiving CZP demonstrated [redacted]
The CZP group in DOSEFLEX experienced [redacted] to health-related quality of life, with improved Short Form (36) Health Survey scores.

Data from the ARTIS registry study were included in the CS and showed significant benefits of CZP treatment in TNFi-experienced patients in DAS28 ($p<0.0001$) and HAQ ($p<0.0001$) at 3 and 6 months following initiation of CZP treatment.

More TNFi-experienced patients in REALISTIC receiving CZP reported an AE (68.1%) than PBO (50.0%). A greater proportion of CZP-treated patients (59.3%) in SWITCH had an AE versus 40.0% on PBO. For TNFi-experienced REALISTIC patients, slightly more CZP-treated patients (7.9%) reported SAEs than those on PBO (5.0%). In DOSEFLEX, CZP treatment groups were more likely to have SAEs (CZP 200 mg every two weeks 7.1%, CZP 400 mg every 4 weeks 2.9%) than on PBO

(0.0%). A greater proportion of TNFi-experienced patients in REALISTIC who received CZP 200 mg reported infections/infestations (29.3%) compared with the PBO group (23.8%).

The results from the Bayesian network meta-analysis (NMA) within the CS indicated that: for population A [redacted]; and for population B [redacted]. No NMA could be performed for population C due to lack of data for the comparators.

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

The ERG was satisfied that the searches for clinical effectiveness evidence reported in the CS were likely to have identified all relevant published RCT evidence. The eligibility criteria applied in the selection of evidence for the clinical effectiveness review were considered by the ERG to be reasonable and generally consistent with the decision problem as outlined in the final NICE scope. The clinical advisors to the ERG did not highlight any additional relevant RCTs that should have been included in the CS. A CZP RCT by Kang *et al.* (2012) was identified by the ERG and clarification sought from the company as to why it was not included in the CS. The company responded that the Kang trial was not included in the CS because only low numbers of patients in the trial were TNFi-experienced. However, the ERG noted that two CZP RCTs were included in the CS that had low numbers of TNFi-experienced patients (J-RAPID and HIKARI) and therefore the ERG considered that additional justification should have been provided by the company to support their decision to exclude the Kang trial.

The quality of the included CZP RCTs and ARTIS non-randomised study were assessed using well established and recognised criteria. Data for radiological progression and joint damage were not presented in the CS, however, data on inhibition of joint structural damage were available in the published articles for both J-RAPID and HIKARI. Extra-articular manifestations of disease were not included in the CS. Study and patient characteristics for included CZP trials were clearly described in a narrative summary alongside clinical and safety data. However, *p*-values were frequently unreported in the CS and therefore the ERG requested that these be provided by the company where available. Classical meta-analyses were performed for CZP used in combination with MTX and for CZP as monotherapy. Classical meta-analyses were performed separately for the outcomes of ACR20/50/70; EULAR response; and DAS28(Erythrocyte Sedimentation Rate (ESR)) remission at 3 months. No meta-analysis was performed for outcomes at 6 months due to data unavailability. Both fixed effects (Mantel-Haenszel) and random effects (DerSimonian and Laird) models were used. Heterogeneity between trials was investigated using I^2 values. The ERG noted that it is generally recommended that

treatment from the REALISTIC study was assumed by the company to represent the evidence for the target population; however, the company only used the "no response" rates from the REALISTIC study and used evidence from all other studies to estimate the response rates for other ACR and EULAR response rates. The company generated approximate estimates of absolute probabilities being in each ACR and EULAR category by not using the results of the NMA appropriately to estimate them.

The ERG believes that the treatment sequences compared for Population A are inappropriate because they include TOC + MTX followed by ABA + MTX after RTX + MTX. Clinical experts consulted by the ERG claimed that usually TOC + MTX or ABA + MTX were provided, but not both.

The ERG notes that the company did not identify evidence on the efficacy of IFX, ADA and ETA in combination with MTX in patients with inadequate response to a TNFi. Similarly, the ERG notes that the company did not identify evidence on the efficacy of TOC, ADA and ETA monotherapies in patients with inadequate response to a TNFi. The ERG believes that the assumptions made by the company to overcome the lack of evidence for its base case analysis did not evaluate fully the uncertainty and that therefore the results of the base case analysis should be interpreted with caution.

The ERG believes that the methodology for modelling first and subsequent treatments is limited and can result in implausible sequences when comparing elongated sequences, such as the intervention sequence in Population A with shorter sequences (the comparator in Population A).

The ERG has concerns regarding the modelling of the efficacy of subsequent treatments due to the lack of evidence on treatment efficacy in patients with an inadequate response to a previous TNFi.

The ERG notes that the company assumed the same treatment duration for all bDMARDs for its base case analysis, despite evidence suggesting different treatment durations for different bDMARDs. The ERG notes that the company identified treatment duration as a parameter with a large impact on the ICER (especially in Population A) in one of their scenario analyses.

The ERG notes that the company used a rather simple approach to map changes in HAQ score to changes in EQ-5D utility and better approaches exist to capture the non-linearity of the relationship between HAQ score and EQ-5D.

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

1.6.1 Strengths

The ERG were satisfied that the key RCTs for the efficacy of CZP in patients with an inadequate response or intolerance to TNFis were included in the CS. The included CZP RCTs were considered by both the company and the ERG to be of good quality.

The model used appears conceptually appropriate with only a few minor implementation errors, some of which were fixed during the clarification process. It contained the functionality to assess the impact of changing parameters and relevant structural uncertainties on the ICER. A number of built-in alternative scenarios were included.

1.6.2 Weaknesses and areas of uncertainty

The ERG considers that additional justification for the omission of the Kang trial from the CS should have been provided by the company. Whilst the REALISTIC trial contained the largest TNFi-experienced population, some of the findings for individual trials in the clinical effectiveness review were based on small sample sizes.

The company performed two sets of NMAs: 1) analyses of the individual categories of ACR and EULAR, 2) separate analyses of the ACR and EULAR data by assuming that the ACR and EULAR categories formed single sets of ordinal data, respectively. In both cases, the company's preferred results assumed that the treatment effects were constant across studies ignoring potential heterogeneity. Absolute probabilities of being in each ACR and EULAR category ignored the underlying joint distribution between parameters generated by the second set of NMAs, underestimated uncertainty and used evidence from all studies to partition responders between different categories rather than using evidence only from the REALISTIC study.

A further area of uncertainty is the efficacy of ADA, IFX, ETA in combination with MTX and of TOC, ADA and ETA monotherapies in patients who have had an inadequate response to a TNFi.

The company did not identify relevant evidence and assumed that the efficacy of ADA, IFX and ETA in combination with MTX was equal to that of GOL + MTX and that the relative efficacy of TOC, ADA and ETA monotherapies compared with CZP monotherapy would be the same as that when these interventions were used in combination with MTX.

The lack of published evidence on the efficacy of competitor bDMARDs in patients who have had an inadequate response to two or more bDMARDs introduced considerable uncertainty in the model. The company adopted simplifying assumptions in order to model the efficacy of

Estimates of the cost-effectiveness of CZP when the ABA and TOC PASs are taken into consideration are provided in a confidential appendix.

[redacted]

After the failure of the first TNF-inhibitor, TA195[25] recommends RTX in combination with MTX for the treatment of severe active RA. If RTX is contraindicated or withdrawn because of an adverse event (AE), TA195 recommends ABA, ADA, ETA, or IFX in combination with MTX. If MTX is contraindicated or withdrawn because of an AE, TA195 recommends ADA or ETA as monotherapy. TA247[26] recommends TOC as an alternative to TNF-inhibitors in the same circumstances as TA195, that is, after the failure of a TNF-inhibitor in patients with severe active RA, in combination with MTX when RTX is contraindicated or withdrawn and as monotherapy if MTX is contraindicated or withdrawn. In addition, TA247 recommends TOC in combination with MTX in patients in whom TNF-inhibitors and RTX have not worked.

A simplified summary of NICE recommendations on bDMARDs is shown in Figure 1. It defines the sequence of treatments that have received positive guidance for patients with a DAS28 score of >5.1. In summary, the typical route would be intensive cDMARDs followed by a bDMARD, followed by RTX plus MTX, then TOC before returning to cDMARDs.

NICE criteria for continuing treatment

NICE TA375[24] states that for patients to continue treatment with their first bDMARD treatment they must maintain at least a moderate EULAR response. TA195 states that bDMARD treatment after the failure of a TNFi should be continued only if there is an adequate response (defined as an improvement in the DAS28 score of at least 1.2 points) at initiation of treatment and as long as this adequate response is maintained. If the criterion of having at least a moderate EULAR response at six months has not been met, then treatment should be stopped and the next intervention in the sequence initiated.

every 4 weeks (Q4W) can be considered. MTX should be continued during treatment with CZP where appropriate. The price of a 200mg syringe prefilled with CZP is £357.50.[30] The company has agreed a PAS with the Department of Health where the first 12 weeks of therapy, consisting of 10 syringes pre-loaded with 200mg of CZP are provided to the NHS free of charge.

1.4 Comparators

Four comparators were defined in the final NICE scope, three of which were considered by the company. In population A, the intervention is added into a treatment sequence before RTX + MTX forming a comparison of an elongated sequence compared with a standard sequence. Other potential sequences, such as replacing RTX + MTX with CZP + MTX, or comparing the elongated sequence with an equally long sequence with RTX + MTX before CZP + MTX were not considered, and thus a fully incremental analysis of all sequences has not be undertaken within the CS.

For patients for whom RTX is contraindicated or has been withdrawn due to an AE, population B, only the first line of therapy is assumed to differ being either the intervention + MTX or one of the comparators named in the final scope (ABA, ADA, ETA, GOL, IFX, TOC) + MTX.

For patients for whom RTX therapy cannot be given because MTX is contraindicated or withdrawn, population C, only the first line of therapy is assumed to differ being either the intervention or one of the comparators named in the final scope (ADA, ETA, TOC).

The company did not incorporate the fourth comparator listed in the NICE final scope[27] which was best supportive care. The reason provided for the company was that this 'does not reflect the current NICE recommendations for TNFis (NICE Pathways for Drug Treatment for RA (26th March 2015); NICE commissioning algorithm (May 2013)). Furthermore, limited evidence supports the evaluation of CZP within this patient group.'

Further details on the comparators assumed in the company's health economic analysis are provided in Sections 5.2.3 and 5.2.8.

1.5 Outcomes

The majority of outcomes reported in the final NICE scope[27] have been included in the CS. Three defined outcome measures have, however, been excluded. These are: joint damage; radiological progression; and extra-articular manifestations of the disease.

4.2.3 Participant flow and numbers

Five CZP RCTs included mixed populations of TNFi-naïve and TNF-experienced patients (with the exception being SWITCH, which included solely TNFi-experienced subjects). REALISTIC was the largest study, with 851 (320 TNFi-experienced) and 212 (80 TNFi-experienced) patients randomised to CZP and PBO, respectively.

In the REALISTIC trial, a high proportion (≥ 85%) of patients in both CZP and PBO treatment arms completed the double-blind controlled phase to 12 weeks, with occurrence of AEs and lack of efficacy being the most common specified reasons for discontinuation. The majority (≥ 80%) of randomised patients also completed the open-label period. The majority of patients (≥ 78%) randomised at week 18 of DOSEFLEX also completed the double-blind phase to week 34 of the study. AEs were the most common reason for withdrawal in the CZP treatment arms, whilst (as might be anticipated) loss of efficacy contributed to most drop-outs in the PBO arm. It is worth noting that DOSEFLEX included an open-label CZP run-in phase that resulted in a considerable number of patients (93/333 [27.9%]) who entered the run-in phase dropping out due to lack of efficacy and not being randomised into the double-blind period. It is also interesting that, whilst a large proportion of patients completed to week 12 of PREDICT (≥ 80% in both CZP groups), only approximately 50% remained at week 52, with lack of efficacy being cited as the most frequent reason for CZP discontinuation. Over 90% of randomised patients in the SWITCH trial completed 12 weeks of the study. Whilst the completion rate of randomised patients to 24 weeks was good (≥ 70%) in the CZP arms of J-RAPID and HIKARI, only 32.5% and 15.8% of PBO group patients completed (with most withdrawing at week 16 due to not achieving an ACR 20 at weeks 12 and 14, as per study design).

In most CZP RCTs (REALISTIC, PREDICT, DOSEFLEX, J-RAPID and HIKARI), missing data were imputed using the last observation carried forward (LOCF) or non-responder imputation (NRI) methods (for continuous and categorical outcome measures respectively). Data imputation methods used in SWITCH were not reported by Schiff *et al.*[37] and no further details could be identified in the trial record on ClinicalTrials.gov.

4.2.5 Summary of clinical effectiveness results for CZP

The statistical significance of comparisons made between CZP and comparator treatment arms in the TNFi-experienced trial populations were frequently not reported in the CS. The ERG requested that the company provide *p*-values where these were not reported in the CS. In response (see clarification response, Question A36[28]), the company performed *post hoc* analyses for key outcomes and provided a series of *p*-values. The company noted that analyses were exploratory (i.e. nominal *p*-values) so should be treated with caution (particularly in cases where *p*-values were based on sample sizes ≤ 15% of study population) and that no conclusions can be made on the statistical significance of comparisons. The *p*-values provided by the company in response to the ERG's clarification request have been added to the ERG report data and are marked with a symbol (†).

4.2.5.1 Disease activity

Disease activity was reported in the CS in terms of ACR and EULAR responses, DAS28 and CDAI. The summary of ACR and EULAR response data, as key disease activity outcomes, has been prioritised in this ERG report. DAS28 (REALISTIC [CS Section 4.7.1.2.2], DOSEFLEX [CS Section 4.7.2.2.2], PREDICT [CS Section 4.7.3.2.2], SWITCH [CS Section 4.7.4.1.2], J-RAPID [CS Section 4.7.5.1.2], HIKARI [CS Section 4.7.5.2.2]) and CDAI (REALISTIC [CS Section 4.7.1.2.2], DOSEFLEX [CS Section 4.7.2.2.2], PREDICT [CS Section 4.7.3.2.2]) data were included in the CS but are not summarised in the ERG report.

4.2.5.2 ACR response

ACR response data were available from all included CZP RCTs (PREDICT reported modified ACR [mACR]) and were collated by the ERG in Table 10. The modified ACR was described in Appendix 8.6.2 of the CS and was reported to differ from the standard ACR in two aspects. Firstly, tender and swollen joints were assessed in 28 joints (used in the DAS28 assessment) [redacted] Secondly, patients' assessment of physical function, global health and pain [redacted] ACR data were provided in the CS for overall trial populations and subgroups of TNFi-experienced patients. Results for TNFi-experienced subjects only are summarised in this section.

Table 10: ACR response rates in included CZP RCTs[*]

Trial name	Treatment Group	Treatment arms for which data extraction performed (n)	Assessment time point	% achieving ACR20 response	% achieving ACR50 response	% achieving ACR70 response
REALISTIC	Overall population	PBO (n=212)	Week 2	[redacted]	[redacted]	[redacted]
		CZP 200 mg Q2W (n=851)	Week 2	[redacted]	[redacted]	[redacted]
		PBO (n=212)	Week 6	[redacted]	[redacted]	[redacted]
		CZP 200 mg Q2W (n=851)	Week 6	[redacted]	[redacted]	[redacted]
		PBO (n=212)	Week 12	[redacted](25.9%)	[redacted](9.9%)	[redacted](2.8%)
		CZP 200 mg Q2W (n=851)	Week 12	[redacted](51.1%) p <0.001	[redacted](26.6%) p <0.001	[redacted](12.9%) p <0.001
		CZP 200 mg OLE (n=770)	Week 12	[redacted]	[redacted]	[redacted]
		CZP 200 mg OLE (n=770)	Week 28	[redacted]	[redacted]	[redacted]
	Overall population, (NRI), CZP monotherapy	PBO (n=69)	Week 12	[redacted]	[redacted]	[redacted]
		CZP 200 mg Q2W (n=262)	Week 12	[redacted]	[redacted]	[redacted]
		CZP 200 mg OLE (n=237)	Week 12	[redacted]	[redacted]	[redacted]
		CZP 200 mg OLE (n=237)	Week 28	[redacted]	[redacted]	[redacted]
	Overall population, (NRI), CZP+MTX	PBO (n=143)	Week 12	[redacted]	[redacted]	[redacted]
		CZP 200 mg Q2W (n=589)	Week 12	[redacted]	[redacted]	[redacted]
		CZP 200 mg OLE (n=533)	Week 12	[redacted]	[redacted]	[redacted]
		CZP 200 mg OLE (n=533)	Week 28	[redacted]	[redacted]	[redacted]
REALISTIC	TNFi-experienced	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	TNFi-experienced (NRI), CZP monotherapy	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	TNFi-experienced (NRI), CZP+MTX	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
DOSEFLEX	Overall population	PBO+MTX (n=69)	Week 4	[redacted]	[redacted]	[redacted]
		CZP 200 mg Q2W+MTX (n=70)	Week 4	[redacted]	[redacted]	[redacted]
		CZP 400 mg +MTX Q4W (n=69)	Week 4	[redacted]	[redacted]	[redacted]
		PBO+MTX (n=69)	Week 12	[redacted]	[redacted]	[redacted]

Trial name	Treatment Group	Treatment arms for which data extraction performed (n)	Assessment time point	% achieving ACR20 response	% achieving ACR50 response	% achieving ACR70 response
		CZP 200 mg Q2W+MTX (n=70)	Week 12	[redacted]	[redacted]	[redacted]
		CZP 400 mg +MTX Q4W (n=69)	Week 12			
		PBO+MTX (n=69)	Week 16			
		CZP 200 mg +MTX Q2W (n=70)	Week 16			
		CZP 400 mg +MTX Q4W (n=69)	Week 16			
		PBO+MTX (n=69)	Week 24			
		CZP 200 mg +MTX Q2W (n=70)	Week 24			
		CZP 400 mg +MTX Q4W (n=69)	Week 24			
		PBO+MTX (n=69)	Week 34	(44.9%)	(30.4%)	(15.9%)
		CZP 200 mg +MTX Q2W (n=70)	Week 34	(67.1%)	(50.0%)	(30.0%)
		CZP 400 mg +MTX Q4W (n=69)	Week 34	(65.2%)	(52.2%)	(37.7%)
	TNFi-experienced (NRI)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	TNFi-experienced (LOCF)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
SWITCH[¥]	TNFi-experienced	PBO Q2W + cDMARDs (n=10)	Week 12	0 (0%)	0 (0%)	0 (0%)
		CZP 200 mg Q2W+cDMARDS (n=27)	Week 12	17 (61.5%) $p<0.005$	5 (19.0%)	1 (3.5%)
		PBO Q2W + cDMARDs (n=8)	Week 24	5 (62.5%) (p value NR)	3 (37.5%) (p value NR)	NR
		CZP 200 mg Q2W+cDMARDS (n=22)	Week 24	12 (54.5%) (p value NR)	6 (27.3%) (p value NR)	NR
HIKARI	Overall population	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	TNFi-	PBO (n=10)	Week 12	[redacted]	[redacted]	[redacted]

Trial name	Treatment Group	Treatment arms for which data extraction performed (n)	Assessment time point	% achieving ACR20 response	% achieving ACR50 response	% achieving ACR70 response
	experienced	CZP 200 mg Q2W (n=6)	Week 12	[redacted]	[redacted]	[redacted]
		PBO (n=10)	Week 24	[redacted]	[redacted]	[redacted]
		CZP 200 mg Q2W (n=6)	Week 24	[redacted]	[redacted]	[redacted]
J-RAPID	TNFi-experienced	PBO +MTX (n=15)	Week 12	[redacted]	[redacted]	[redacted]
		CZP 200 mg Q2W +MTX (n=11)	Week 12	[redacted]	[redacted]	[redacted]
		PBO +MTX (n=15)	Week 24	[redacted]	[redacted]	[redacted]
		CZP 200 mg Q2W +MTX (n=11)	Week 24	[redacted]	[redacted]	[redacted]

* The PREDICT trial measured mACR instead of standard ACR (data are summarised in supporting ERG report text)

¥ SWITCH 24-week data were sourced from Clinicaltrials.gov. Efficacy denominators differ from N randomised (reason unclear) and are as reported in the source material

† *p*-values provided in company's clarification response

Figure 2: PREDICT study: kinetics of mACR20/50/70 response rates in TNFi experienced population during 52-week double-blind phase (NRI) (reproduced from CS Figure 35)

mACR responses among TNF-experienced patients receiving CZP monotherapy and MTX combination treatment in PREDICT were presented in Appendix 8.9.1.1 of the CS

Patients in the SWITCH trial who were TNFi-experienced and treated with CZP + cDMARDs had more favourable ACR20, 50 and 70 responses by week 12 than those in the PBO group (ACR20 $p<0.005$). The company clarified that ACR50/70 data at week 12 and response rates at week 24 were estimated using graph-reading software. However, the ERG identified that 24-week data for SWITCH were available on ClinicalTrials.gov. The denominators used in the reported analysis on ClinicalTrials.gov differed from the randomised total for unspecified reasons. Interestingly, analyses reported on ClinicalTrials.gov indicated that the proportions of ACR20 and ACR50 responders at week 24 (end of the 12-week OLE period, during which all patients received CZP) were lower in patients initially treated with CZP (54.5% and 27.3%) compared with patients initially treated with PBO (62.5% and 37.5%) (but analyses were based on relatively small numbers of subjects).

In their clarification response (see clarification response, Question A36[28]), the company stated [redacted] of CZP 200 mg Q2W patients had a good or moderate EULAR response versus [redacted] PBO patients [redacted] TNFi-experienced patients receiving CZP + MTX in REALISTIC were [redacted] to achieve good or moderate EULAR responses [redacted] at week 12 than those on PBO [redacted] CZP + MTX-treated patients were more [redacted] than PBO subjects [redacted] to have a good or moderate EULAR response at 12 weeks [redacted])[†]. The proportion of patients achieving a good EULAR response [redacted] for patients on CZP + MTX than CZP monotherapy [redacted] In their clarification response (see clarification response, Question A36[28]), the company stated that in the TNFi-experienced population, CZP 200 mg Q2W monotherapy-treated patients [redacted] to reach a good or moderate EULAR response at week 12 than PBO patients [redacted].

In DOSEFLEX, [redacted] of TNFi-experienced patients in the study arms achieved EULAR good or moderate response rates at [redacted]). However, by week 34, TNFi-experienced patients treated with CZP 200 mg Q2W + MTX [redacted] and CZP 400 mg Q4W + MTX [redacted] were much more [redacted] to reach a good EULAR response than those receiving PBO + MTX [redacted] In their clarification response (see clarification response, question A36[28]), the company reported that CZP-treated patients were more [redacted][†] to be EULAR good responders (with non-response defined as EULAR moderate or no response) compared with PBO group patients. The company provided in their clarification response pooled data from DOSEFLEX AND DOSEFLEX II for EULAR response to week 42 in the TNFi-experienced population randomised to CZP + MTX from week 18. The method of data pooling was not described. The proportions of patients achieving good or moderate responses [redacted]

The proportions of TNFi-experienced patients treated with CZP reaching moderate or good EULAR responses in the PREDICT trial were [redacted] maintained between weeks 12 and 52[redacted]

TNFi-experienced patients receiving CZP + MTX in the J-RAPID trial were more [redacted] to reach moderate [redacted] or good [redacted] EULAR responses at week 12 than PBO group subjects [redacted]

In the HIKARI trial, [redacted] of CZP-treated TNFi-experienced patients achieved good [redacted] or moderate [redacted] EULAR responses at week 12 compared with PBO [redacted]

CZP benefits were maintained at 24 weeks for good EULAR response [redacted] between groups for moderate EULAR response [redacted]

EULAR good or moderate response data at weeks 12 and 24 were available for SWITCH in ClinicalTrials.gov. The reason for the different denominators in the analyses for weeks 12 and 24 was unclear. Whilst CZP-treated patients demonstrated a much more favourable EULAR response at week 12, results at week 24 were roughly comparable between CZP and PBO treatment groups; *p*-values were not reported.

Table 14: EULAR good or moderate response in SWITCH (source: Clinicaltrials.gov)

Treatment arm	Time of assessment	Number (%) of patients achieving a EULAR good or moderate response*
CZP 200 mg Q2W + DMARDs	Week 12	17/26 (65.4%)
PBO Q2W + DMARDs	Week 12	0/9 (0.0%)
CZP 200 mg Q2W + DMARDs	Week 24	17/22 (77.3%)
PBO Q2W + DMARDs	Week 24	6/8 (75.0%)

* *p*-values not reported

EULAR data at 3 months were meta-analysed to compare the effects of CZP (both in combination with MTX and as monotherapy) with PBO. The forest plots indicated [redacted] favourable effects for the CZP groups in terms of EULAR response, although this effect appears [redacted] for CZP as monotherapy and these results were [redacted]

4.2.7 Physical function

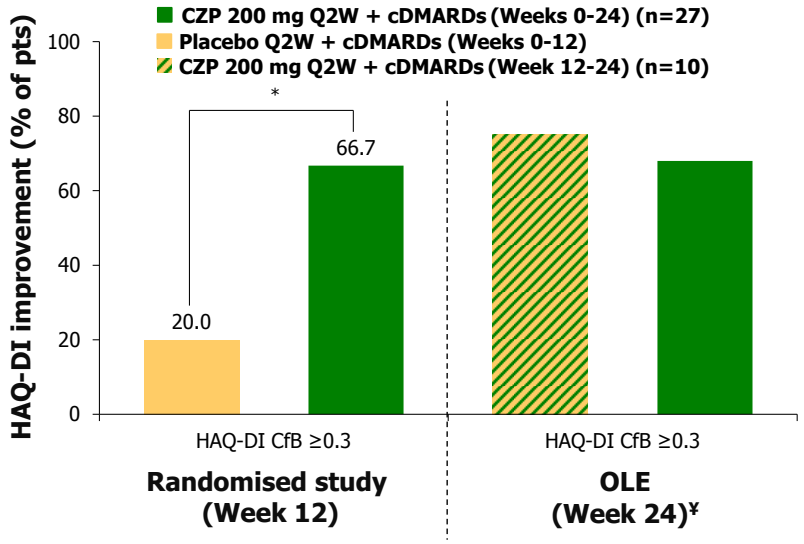
Four of the included CZP RCTs reported physical function in terms of Health Assessment Questionnaire Disability Index (HAQ-DI) scores (or MD-HAQ/M-HAQ in the case of PREDICT). Whilst HAQ-DI outcome data were collected at weeks 12 and 24 in the J-RAPID[32] and HIKARI[33] studies, these data were not presented in the CS.In the TNFi-experienced patients in REALISTIC, the mean HAQ-DI changes in both CZP [redacted] and PBO [redacted] treatment groups showed [redacted] from baseline to week 12 [redacted] The CS states that for HAQ-DI mean change there was [redacted] interaction between treatment and TNFi experience [redacted]

Figure 5: REALISTIC study: HAQ-DI to Week 12 for TNFi experienced population during 12 week double-blind phase (OLE scores in inset table) (LOCF) (reproduced from CS Figure 24)

It was also reported in the CS that HAQ-DI scores within the REALISTIC RCT were [redacted] For TNFi-experienced patients treated with CZP+MTX, the HAQ-DI mean change from [redacted]

Both study groups in the DOSEFLEX trial showed [redacted] HAQ-DI during the 16 week open label CZP run-in period [redacted] However, differences between groups were evident following randomisation into the double-blind phase, with [redacted]

Figure 8: SWITCH study: HAQ-DI improvement for TNFi experienced (overall) population at Weeks 12 and 24 of OLE phase (reproduced from CS Figure 41)



*p=0.046 CZP vs PBO. ¥Data for Week 24 have been re-drawn from the manuscript. Data point values are not available. Figure adapted from Schiff *et al.* 2014[37]

4.2.8 Joint damage/radiological progression

No data for the outcomes of joint damage or radiological progression were included in the CS. However, data on inhibition of joint structural damage were reported in the published articles for both J-RAPID and HIKARI. Both trials included radiographic assessments at baseline and week 24 or at discontinuation using modified Total Sharp Score (mTSS). Due to time constraints these data have not been extracted by the ERG.

4.2.9 Pain

Pain was listed as a secondary outcome in the REALISTIC, DOSEFLEX, J-RAPID and HIKARI RCTs.

Bodily pain as a domain of the SF-36 was reported for the DOSEFLEX study [redacted] Both PBO and CZP (200 mg Q2W and 400 mg every 4 weeks Q4W) treatment groups experienced [redacted] in bodily pain score [redacted] During the double-blind period [redacted] PBO group patients had [redacted] SF-36 bodily pain score, whilst scores were [redacted] receiving CZP.

Data for pain as a component of the ACR response were presented for TNFi-experienced patients in REALISTIC (Appendix 8.7.3.1 of the CS and updated in clarification response Question A36[28]) and DOSEFLEX (Appendix 8.8.3.1 of the CS).

Subjects receiving CZP 200 mg Q2W in REALISTIC demonstrated a [redacted] mean reduction in patient's assessment of arthritis pain score from baseline to week 12 [redacted] than placebo group patients [redacted] Benefits were maintained between weeks 12 and 28 in patients receiving CZP 200 mg Q2W in the OLE [redacted] in patient's assessment of arthritis pain score from baseline to week 12 were evident in TNFi-experienced patients in REALISTIC in both CZP monotherapy [redacted] vs. PBO [redacted] combination with MTX subgroups [redacted]

In the open-label CZP run-in period to week 16 of the DOSEFLEX study [redacted] in patient's assessment of arthritis pain as an ACR component were [redacted] for all treatment groups. However, during the randomised controlled period (week 18 to week 34), PBO plus MTX group patients experienced [redacted] whilst pain scores were [redacted]

Data on patient's assessment of arthritis pain were reported in the trial publications for J-RAPID (Yamamoto 2014a[32]) and HIKARI (Yamamoto 2014b[33]) but were not presented in the CS.

4.2.10 Mortality

Data relating to deaths among CZP RCT participants are presented in the safety section of this report.

4.2.11 Fatigue

The impact of CZP on fatigue was assessed in the REALISTIC and DOSEFLEX studies.

In the REALISTIC RCT patients were assessed using the Fatigue Assessment Scale. Subjects were required to answer the following question: "Please rate your fatigue (weariness, tiredness) during the past 7 days, on a scale of 0 to 10" where 0 is 'No Fatigue' and 10 is 'Fatigue as bad as you can imagine'. Clinically meaningful [redacted] in fatigue (defined as ≥1-point improvement) were observed in both CZP [redacted] and PBO [redacted] groups at week 12. Benefits of CZP treatment were maintained in the OLE [redacted]

Figure 9: REALISTIC study: Patient's Assessment of Fatigue in TNFi experienced population during 12 week double-blind phase (OLE scores in inset table) (LOCF) (reproduced from CS Figure 22)

Fatigue data were broken down by TNFi-experienced monotherapy and combination with MTX subgroups (Appendix 8.7.1.7 of the CS, see clarification response Question A36[28]) fatigue was observed at week 12 for CZP-treated patients compared with PBO in the CZP monotherapy subgroup. The difference in fatigue between treatment arms was CZP combination subgroup for PBO group subjects.

The DOSEFLEX trial utilised the 10-point Fatigue Assessment Scale. Fatigue improved to week 16 with CZP during the open-label run-in in all study groups Whilst fatigue improvements were maintained to week 34 of the randomised controlled period in both CZP treatment groups, the PBO group experienced a worsening of fatigue.

4.2.16.2 Serious adverse events

The proportions of patients experiencing SAEs in the REALISTIC trial were similar (CZP 200 mg Q2W 6.1%; PBO 5.7%). Specific reported reasons for SAEs in the DOSEFLEX CZP groups included serious infections (CZP 200 mg Q2W 4.3%), cardiac disorders (CZP 200 mg Q2W, 1.4%), and musculoskeletal/connective tissue disorders (CZP 200 mg Q2W 2.9%, CZP 400 mg Q4W, 1.4%). For TNFi-experienced REALISTIC patients, slightly more patients on CZP 200 mg Q2W (7.9%) reported SAEs than those on PBO (5.0%) (reasons for SAEs unspecified). Both CZP treatment groups in the DOSEFLEX randomised controlled period had greater proportions of SAEs (CZP 200 mg Q2W 7.1%, CZP 400 mg Q4W 2.9%) than those in the PBO group (0.0%). No serious adverse events (SAEs) were reported for SWITCH participants.

4.2.16.3 Adverse events leading to withdrawal/discontinuation of treatment

During the first 12 weeks of REALISTIC, 3.8% of PBO group subjects had AEs leading to withdrawal, with 4.7% of cases in the CZP 200 mg Q2W arm. Of the TNFi-experienced subjects in REALISTIC, slightly more CZP 200 mg Q2W-treated patients (5.4%) had AEs leading to withdrawal than PBO group participants (2.5%). However, more patients in the REALISTIC CZP treatment arms had AEs that led to permanent discontinuation (CZP 200 mg Q2W 5.7%, CZP 400 mg Q4W 1.4%) compared with none in the PBO group. In the randomised controlled phase of DOSEFLEX, patients in all groups experienced AEs leading to withdrawal (CZP 200 mg Q2W 17.1%; CZP 400 mg Q4W 8.7%; PBO 11.6%).

4.2.16.4 Infections

Infections were more common for patients receiving CZP 200 mg Q2W (29.0%) in REALISTIC compared with PBO (23.0%). A greater proportion of TNFi-experienced patients in REALISTIC who were treated with CZP 200 mg Q2W had infections/infestations (29.3%) versus the PBO group (23.8%). The proportions experiencing infections in the randomised controlled period of DOSEFLEX were similar across CZP (CZP 200 mg Q2W 28.6%, CZP 400 mg Q4W 36.2%) and PBO (34.8%) treatment arms. Infections were not reported in the SWITCH trial.

4.2.16.5 Injection site reactions

Injection site reactions appeared more common for patients receiving CZP compared with PBO to the end of the 12 week double-blind phase of REALISTIC (CS Table 52). Among the overall population in REALISTIC, a greater proportion of CZP patients (CZP 200 mg Q2W→CZP 200 mg Q2W, 4.9%) had injection site reactions compared with those who switched from PBO to CZP 200 mg Q2W (1.8%).

Table 17: Adverse events in overall study populations for CZP RCTs[¥] (collated by ERG)

AEs	DOSEFLEX			REALISTIC		SWITCH	
	CZP 200 mg Q2W (n=70)	**CZP 400 mg Q4W (n=70)**	**PBO (n=69)**	**CZP 200 mg Q2W (n=846)**	**PBO (n=209)**	**CZP 200 mg Q2W + cDMARDs → CZP 200 mg Q2W + cDMARDs (n=27)**	**PBO + cDMARDs → CZP 200 mg Q2W + cDMARDs (n=10)***
AEs, incidence rate/100 PY (n, patient %)						**AEs, n (patient %)**	
Any AEs	312.1 (44, 62.9)	299.9 (42, 60.9)	323.6 (43, 62.3)	522.1 (571, 67.5)	483.2 (129, 61.7)	16 (59.3)	4 (40.0)
Mild	-	-	-	248 (29.3)	56 (26.8)	7 (25.9)	3 (30.0)
Moderate	-	-	-	257 (30.4)	58 (27.8)	9 (33.3)	1 (10)
Severe	-	-	-	66 (7.8)	15 (7.2)	0 (0)	0 (0)
Infection and infestations	104.9 (20, 28.6)	132.4 (25, 36.2)	136.2 (24, 34.8)	143.9 (245, 29.0)	112.5 (48, 23.0)	-	-
Lower respiratory tract infections	-	-	-	3.5 (7, 0.8)	2.1 (1, 0.5)	-	-
Upper respiratory tract infections	23 (5, 7.1)	36.2 (8, 11.6)	46.5 (10, 14.5)	59.3 (112, 13.2)	41.5 (19, 9.1)	5 (17.6)	0 (0)
Nasopharyngitis	4.4 (1, 1.4)	4.4 (1, 1.4)	18.4 (4, 5.8)	-	-	-	-
Sinusitis	9 (2, 2.9)	13.1 (3, 4.3)	0 (0)	-	-	-	-
Urinary tract infection	23.1 (5, 7.1)	27.6 (6, 8.7)	33.4 (7, 10.1)	2.5 (5, 0.6)	4.2 (2, 1.0)	-	-
Ear infections	0 (0)	13.3 (3, 4.3)	0 (0)	-	-	-	-
Streptococcal infections	-	-	-	0 (0)	2.1 (1, 0.5)	-	-
Musculoskeletal/connective tissue disorders	37.6 (8, 11.4)	51.4 (11, 15.9)	64.2 (13, 18.8)	-	-	-	-
Arthralgia	4.5 (1, 1.4)	22.5 (5, 7.2)	8.9 (2, 2.9)	-	-	-	-
Back pain	13.5 (3, 4.3)	0 (0)	4.4 (1, 1.4)	-	-	-	-
RA aggravation	4.4 (1, 1.4)	8.9 (2, 2.9)	27.7 (6, 8.7)	-	-	-	-
Pain in extremity	8.9 (2, 2.9)	0 (0)	13.5 (3, 4.3)	-	-	-	-
Nervous system disorders	22.8 (5, 7.1)	17.8 (4, 5.8)	4.4 (1, 1.4)	-	-	-	-
Dizziness	13.5 (3, 4.3)	0 (0)	4.4 (1, 1.4)	-	-	-	-
Headache	9 (2, 2.9)	0 (0)	0(0)	24.2 (47, 5.6)	23.5 (11, 5.3)	2 (7.4)	0 (0)

overestimate the time between consecutive courses of rituximab. However, based on clinical specialists' advice, the Committee considered it was unlikely that the mean retreatment interval would be as low as 6 months. The company assumed a 6-month retreatment interval for the base case analysis but explored the impact on the ICER of a 9-month retreatment interval in a scenario analysis as assumed in TA375.[24]

The company's model uses a mapping from HAQ score to EQ-5D. In the base case, a linear mapping attributed to Brennan *et al.*[70] is used which maps changes in HAQ to changes in EQ-5D by multiplying them by -0.2102. The company undertook a scenario analysis using the algorithm proposed by Hernández Alava *et al.* [65] but due to issues in its implementation (explained in Section 5.3) the ERG decided not to replicate these results in this report.

In order to estimate the change in utility from baseline on initial response to the first therapy considered in the analysis (i.e., second bDMARD), the company fitted a linear regression model to the data from the PREDICT study. A scenario analysis explored the impact on the ICER of estimating this utility using the same mapping from HAQ to EQ-5D as in the rest of the model.

The company assumed within the base case analysis that the patients had reached the utility gain associated with their 6-month EULAR response by week 6. A scenario analysis undertaken by the company explores the impact of assuming a percentage of the utility gain is achieved at week 6 with the remaining utility gained at a constant rate between week 6 and month 6.

The base case analysis assumes that the time to treatment discontinuation for the first treatment is independent of bDMARD (i.e. that the times are equal for all biologics). The company explored the impact of assuming different values for the scale parameter of the Weibull distribution for TNFi (0.3003) and non-TNFi treatments (0.2208), allowing these classes to have differential time to discontinuation.

The CS also includes a scenario analysis that the company claims uses a societal perspective. The only difference with the base case analysis is the inclusion of indirect costs associated with patients' HAQ scores. Indirect costs per HAQ band were taken from a paper reporting costs from the Early RA Study (ERAS).[79] A breakdown of the indirect costs per HAQ score band is provided in Table 48.

Other scenario analyses undertaken by the company assumed: patients on cDMARDs and palliative care would experience no HAQ progression; no vial wastage; and removing the maximum value of 2.76 for HAQ.

not consistently applied to other variables, such as the HAQ score, and outcomes that depend on the HAQ score, such as mortality and HAQ related costs. This slight inaccuracy is likely to underestimate the benefits of the most efficacious bDMARDs, and therefore overestimate the ICER of CZP compared with ABA and the rest of the TNFis and also overestimate the ICER of TOC compared with CZP. However, the ERG notes that the impact of this inaccuracy on the ICERs is likely to be negligible.

(17) Exclusion of adverse events

Adverse events were excluded from the economic analysis, claiming that there was no meaningful difference in the toxicity or risks of AEs between alternative bDMARDs. However, given that the company's model assumes treatment efficacy affects mortality, AEs could be more predominant in patients that live longer. However, the ERG notes that the impact of excluding AEs from the economic analysis is likely to be negligible based on sensitivity analyses performed in TA375.[24]

(18) Inaccuracies in number of doses per cycle

The ERG notes that there is a slight inconsistency in the number of doses assumed for the first six months of ABA (IV) and IFX therapy compared with the subsequent six-month cycles of these drugs or with drugs whose dosing frequency is not a divisor of the cycle length, such as TOC (IV). Even if for ABA (IV) and IFX 8 and 5 doses are respectively administered during the first six months, only 7.5 and 4.5 should be accounted for in the first cycle. However, the ERG notes that the impact of this inaccuracy on the ICERs is likely to be slight.

(19) Appropriateness of using EQ-5D data from the PREDICT study

The utilities used in the company's model were based on EQ-5D data collected in the PREDICT study. However, the population characteristics used in the model were based on the TNFi experienced population of REALISTIC. The ERG considers that it would have been preferable to use EQ-5D scores collected in REALISTIC and that should there be significant differences between the two populations, the utilities currently used in the model would be biased. Unfortunately, EQ-5D scores were not collected in REALISTIC.

(20) Perceived model errors and other issues surrounding model implementation

The cost of TOC (IV) monotherapy was incorrectly calculated because of an error in the model implementation, which led to erroneous results being reported in the scenario analysis for Population C. The error is that the TOC monotherapy administration costs are linked to those of ABA's method of administration (SC or IV).

In order to evaluate Sequence 2, the ERG had to adapt the company's model to support CZP + MTX as a follow-up treatment. In order to do so, the ERG calculated the probability of no response for CZP + MTX following the same approach as the company used to calculate the probability of no response for ABA + MTX and RTX + MTX: using the treatment effect parameters in the NMA to the trial-specific baseline effects from the RADIATE study. The resulting probability of discontinuation for CZP + MTX was estimated to be 44.6%. The ERG used the SC formulation of TOC instead of IV within the sequences because it was less expensive and was assumed to have the same efficacy.[87]

The results of the deterministic and probabilistic analyses using the ERG base case are shown in Table 58 and Table 59. The ERG notes that CZP after RTX (Sequence 2) dominates CZP before RTX (Sequence 1) and that the currently recommended pathway (Sequence 4) dominates the same sequence if RTX is replaced with CZP (Sequence 3). The ERG notes that these results [redacted] shows the cost-effectiveness acceptability curve (CEAC) for the ERG's base case analysis for Population A. The ERG notes that given the limitations of the model when comparing elongated sequences, the value of the CEAC is limited.

It is noticeable that Sequence 4, which includes only three lines of biologics, dominates Sequence 1, which includes four lines of biologics. The ERG believes that this is due to the different methods used for modelling first treatments compared with follow-up treatments, in particular, the fact that the benefits of the first treatment outweigh those of subsequent treatments. For this reason, the ERG believes the model is not appropriate for making comparisons of sequences which include different numbers of treatments and that the fully incremental analyses reported in Table 58 and Table 59 should be interpreted with caution. However, the conclusion that CZP should not be placed before RTX appears to be robust.

For Population B, the probabilistic ICER of CZP + MTX versus ETA(bio) + MTX is expected to be £12,116 per QALY gained and the probabilistic ICER of TOC(SC) + MTX versus CZP + MTX is expected to be £45,414 per QALY gained. These ICERs are less favourable to CZP + MTX than the company's base case ICERs. However, the probability that CZP + MTX produces more net benefit than its comparators assuming a WTP threshold of £30,000 per QALY gained remains essentially unchanged at 0.96.

For Population C, the probabilistic ICER of CZP monotherapy versus ETA(bio) monotherapy is expected to be £13,784 per QALY gained and the probabilistic of TOC(SC) monotherapy versus CZP monotherapy is expected to be £46,501 per QALY gained. These ICERs are less favourable to CZP monotherapy than the company's base case ICERs. However, the probability that CZP monotherapy produces more net benefit than its comparators assuming a WTP threshold of £30,000 per QALY gained is reduced slightly to 0.96.

Additional analyses undertaken by the ERG using this revised base case model indicate that: excluding J-RAPID for the NMA has little impact on the results of the analyses. In contrast, assuming that ADA, IFX and ETA in combination with MTX have the same efficacy as CZP + MTX (rather than GOL + MTX) leads to ETA biosimilar + MTX dominating CZP + MTX; similarly, assuming ADA and ETA monotherapy have the same efficacy as CZP monotherapy leads to ETA biosimilar monotherapy dominating CZP monotherapy. The ERG notes that even were CZP + MTX dominated by ETA biosimilar + MTX there remains comparators for which it is estimated that CZP + MTX is dominant, such as IFX + MTX and ADA + MTX. The latter two interventions will remain options recommended by NICE for treatment in Populations B and C.

With respect to the company's economic analysis and the ERG's additional exploratory analyses, there remain several potentially important areas of uncertainty:

1. The lack of data on the efficacy of ETA, ADA and IFX in combination with MTX in TNFi-IR patients. There is a similar lack of data on the efficacy of ETA and ADA monotherapy in TNFi-IR patients. Different assumptions for the efficacy of these drugs produced markedly different results. This limitation had already been highlighted by the AC of TA195.[25]
2. The scarcity of data on the efficacy of bDMARDs in general, and TNFis in particular, in patients who have had an inadequate response to two or more bDMARDs. There is also the possibility that there could be reduced efficacy of TNFis following inadequate response to a previous TNFi.

3. The relative efficacies of the bDMARDs are uncertain given the limitations of the NMA within the CS.

[redacted]