tofacitinib at 3 months, there are no placebo-controlled results for 6 months for any of the other relevant endpoints in ORAL Solo. Another recently completed head-to-head RCT including tofacitinib monotherapy versus tofacitinib plus methotrexate or adalimumab plus methotrexate (ORAL Strategy) was presented but only as a preliminary result for the primary endpoint of 50% improvement in the American College of Rheumatology (ACR50) in the CS. This RCT found tofacitinib monotherapy was not shown to be non-inferior in efficacy compared to adalimumab plus methotrexate and tofacitinib plus methotrexate at 6 months whilst tofacitinib plus methotrexate was found to be non-inferior to adalimumab plus methotrexate using ACR50 at 6 months.

A revised summary of safety data for tofacitinib provided by the company following an ERG request showed that the highest incidence rates of adverse events (AEs) were for serious infection events and herpes zoster. Additional data provided by the company indicated bronchitis, pneumonia and all cardiac disorders occurred most commonly in the tofacitinib treatment arms.

Network meta-analyses (NMA) were performed to assess the relative efficacy of tofacitinib compared with the comparators in patients who were inadequate responders (IR) to conventional DMARDs (cDMARD-IR) or to biologic DMARDs (bDMARD–IR) patients with moderate-to-severe RA for EULAR response and change in the Health Assessment Questionnaire disability index (HAQ-DI) at 6 months. For the base case NMA cDMARD-IR population, the odds of achieving a EULAR response were all statistically higher for tofacitinib in combination with methotrexate (tofacitinib plus cDMARD) compared to cDMARD at 6 months. No statistically significant differences were found for tofacitinib plus cDMARD versus bDMARDs plus cDMARD, except for tocilizumab plus cDMARD, which was statistically superior in attaining at least a good EULAR response.

Whilst the odds of all EULAR responses were higher in tofacitinib monotherapy compared to cDMARD, only the effect for a good response was statistically significant. No statistically significant differences were found in tofacitinib versus bDMARDs. Both tofacitinib plus cDMARD and tofacitinib monotherapy were associated with significant reduction in HAQ-DI compared with cDMARD at 6 months.

For the base case NMA bDMARD-IR population, the odds of all EULAR responses were all statistically higher in tofacitinib plus cDMARD compared with cDMARD at 6 months. No statistically significant differences were found for tofacitinib plus cDMARD versus abatacept plus cDMARD. Tofacitinib plus cDMARD was statistically superior compared to golimumab plus cDMARD in attaining both at least a moderate and a good EULAR response; but statistically inferior versus rituximab plus cDMARD, tocilizumab plus cDMARD, non-tumour necrosis factors alpha inhibitors (non-TNFi) plus cDMARD and TNFi plus cDMARD. Tofacitinib in combination with

cDMARD was associated with a significant reduction in HAQ-DI compared with cDMARD at 6 months.

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

The ERG considers the searches for clinical effectiveness evidence reported in the CS to be adequate, and believes the included RCTs of tofacitinib to be relevant to the decision problem. It is noted that one recently published RCT (ORAL Strategy) was stated to be "ongoing" in the CS and the company indicated in the CS that results will be available in early May 2017. Following a request from the ERG, the company provided an updated NMA of clinical effectiveness that included data from ORAL Strategy.

The eligibility criteria applied in the selection of evidence for clinical effectiveness were considered by the ERG to be reasonable and consistent with the decision problem outlined in the final NICE scope. The quality of the included RCTs was assessed using well-established and recognised criteria. Primary endpoints and selected analyses for clinical efficacy were appropriate.

The ERG considers that the company's safety overview lacks transparency due to pooling both combination and monotherapy trials to produce incidence rates; the lack of consistent comparison to the control arms; the lack of NMA of adverse events versus comparators; and the failure to search for and provide a complete, comprehensive and up-to-date overview of all AEs including serious adverse events (SAEs). Clinical advice received by the ERG indicates that a more informative AE profile would describe the relative occurrence of all adverse events versus the control arm. Clinical advice received by the ERG also stresses the importance of monitoring the occurrence of AEs for new classes of drugs, and in turn, the importance of searching and including up-to-date evidence to inform the AE profile for the current assessment of tofacitinib. Whilst the CS did not provide a NMA of adverse events versus comparators, the company did reference a paper that conducted a NMA showing that the incidence of herpes zoster was significantly higher for tofacitinib versus bDMARD comparators.

The ERG believes that the results presented in NMA of clinical effectiveness should be treated with caution, as the ordered categorical EULAR data were dichotomised in the cDMARD-IR population, which ignores the natural ordering and correlations between the EULAR response categories. A fixed effects model was used in all the analyses in the bDMARD-IR population, and EULAR response (moderate response and good response) in the cDMARD-IR population. Heterogeneity is expected and this approach underestimates uncertainty in the treatment effects. For tofacitinib trials with early escape, the results from non-responder imputation without advancement penalty (non-responder imputation only applied for the placebo arm, not the tofacitinib arm) were used in the base case

However, the analyses presented by the company included a number of limitations. First, relevant comparators recommended by NICE were missing from the company's analyses: adalimumab, etanercept, infliximab and certolizumab pegol with concomitant MTX in bDMARD-IR RTX-ineligible patients with severe RA and all relevant comparators in bDMARD-IR MTX-intolerant patients with severe RA. The CS did not identify publications for inclusion of adalimumab, infliximab and certolizumab pegol for these populations. Second, the sequences used in the company's original analyses were not appropriate for a number of reasons: (i) the inclusion of multiple consecutive lines of the same treatment; (ii) the inclusion of bDMARD treatments in points in the pathway not recommended by NICE; and, (iii) the inclusion of three or four post-biologic treatments before palliative care. Thirdly, the company assumed equal efficacy for tofacitinib as monotherapy and in combination with MTX in terms of the probabilities of achieving moderate and good EULAR responses. However, the results of the NMA show that these probabilities are lower for tofacitinib monotherapy compared with tofacitinib with concomitant MTX. Fourth, the company used the results for placebo from the NMA to estimate the efficacy sulfasalazine for the analysis for the cDMARD-IR MTX-intolerant population. The ERG believes this to lead to an underestimation of the efficacy of sulfasalazine. Finally, the company rounded modified HAQ-DI values to the nearest valid HAQ-DI score rather than allowing the valid HAQ-DI score to be sampled based on the continuous HAQ-DI value. The ERG notes that this approach might lead to inaccurate estimations of HAQ-DI scores, as values might be rounded up more often than they are rounded down or *vice versa.* The company corrected the first two issues in the revised model submitted with the clarification responses but did not present a full set of analyses relating to their revised base case.

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

1.6.1 Strengths

The ERG considers the data on clinical effectiveness in the CS to be well-reported and the included trials are of good quality.

The model used appears conceptually appropriate with very few implementation errors, most of which were rectified during the clarification process. The ERG considers that the DES approach taken by the company, which was based on the model used in TA375, was deemed appropriate to represent the disease. The ERG considers the company's analysis of patients with moderate RA that can progress to severe RA and then start with a sequence of bDMARDs to reflect the treatment pathway of these patients better than other previous analyses.

The ERG also notes that the amendments, corrections and different assumptions tested by the ERG do not significantly impact the broad conclusions of the analyses presented in the CS.

1.6.2 Weaknesses and areas of uncertainty

Whilst full data were not available for inclusion into the CS, the ERG believes that the recently published ORAL Strategy trial is also relevant to the decision problem because it has head-to-head evidence at 6 months, demonstrating than tofacitinib monotherapy was not shown to be non-inferior to either adalimumab plus MTX and tofacitinib plus MTX using the primary endpoint of ACR50.

The company focuses its safety profile on whether the AEs were comparable across the tofacitinib treatment arms and whether any new or unexpected safety events have occurred. The ERG considers that a more informative analysis would present all AEs, including SAEs, versus the comparator arm. Additionally the company did not conduct targeted up-to-date literature searches to retrieve evidence for AEs associated with tofacitinib treatment for this appraisal meaning that some relevant analyses of adverse event data for tofacitinib are not included. Pooled analyses of AE data across trials of both tofacitinib monotherapy and tofacitinib in combination with methotrexate are unlikely to provide an accurate reflection of the incidence of adverse event rates from these two treatment regimens, which are noted in sources not referenced in the CS, to be different.

1.7 Summary of exploratory and sensitivity analyses undertaken by the ERG

The ERG undertook exploratory analyses based on the company's revised model. The ERG presented two full sets of analyses: one based on the company's preferred NMA and the other based on the NMA undertaken for the clarification response, which the company denoted 'ERG preferred'. As this is not the ERG's preferred analysis we have renamed this the 'clarification NMA'. All analyses presented in this report have not taken any commercial-in-confidence PASs into consideration.

For cDMARD-IR patients with severe RA who can tolerate MTX, based on the company's NMA, tofacitinib + MTX dominated all of its bDMARD comparators except etanercept biosimilar + MTX. Based on the clarification NMA, tofacitinib + MTX dominated ADA+MTX but was extendedly dominated in the full incremental analysis. For cDMARD-IR patients with severe RA for whom MTX was contraindicated or not tolerated, tofacitinib and tocilizumab monotherapy extendedly dominated ADA and ETN biosimilar regardless of the NMA used. The ICER of tocilizumab compared with tofacitinib was £51,488 and £50,430 per QALY gained using the company's NMA and using the clarification NMA, having removed the constraint that TOF monotherapy had the same efficacy as TOF+MTX, respectively.

In the bDMARD-IR patients with severe RA for whom rituximab was an option, rituximab + MTX dominated tofacitinib + MTX regardless of the NMA used. [redacted]. Replacing

tocilizumab + MTX with tofacitinib + MTX after rituximab + MTX was estimated to result in £67,852 and £90,846 per QALY lost using the company's and the clarification NMA respectively. The ERG notes, however, that the confidential PAS of TCZ was not included in these analyses as recommended to the company by NICE at the decision problem meeting. In the bDMARD-IR patients with severe RA for whom RTX was not an option, tofacitinib + MTX dominated golimumab+MTX regardless of the NMA used, and dominated abatacept + MTX also when using the company's NMA. The ICER of etanercept biosimilar and tocilizumab with MTX compared with tofacitinib + MTX was higher than £30,000 per QALY gained regardless of the NMA used.

Finally, in patients with moderate RA who where cDMARD-IR, the ICER of tofacitinib + MTX compared with MTX was £47,594 and £50,708 per QALY gained using the company's and the clarification NMA respectively.

ACR responses,[23] although EULAR is much more closely aligned to the treatment continuation rules stipulated by NICE for treatment in England. 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[22]

	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 regarded as having low disease activity, those with a DAS28 > 3.2 and ≤ 5.1 are regarded as having moderate disease and >5.1 as having very active disease.[21] Within NICE Technology Appraisal (TA) 375, patients with a DAS28 > 3.2 and ≤ 5.1 were considered as having moderate-to-severe disease whilst those with a DAS28 > 5.1 were denoted as having severe disease.[24]

A widely used measure of patient disability is the Health Assessment Questionnaire (HAQ). The HAQ-DI score is a patient completed disability assessment which has established reliability and validity.[25] HAQ-DI scores range from zero to three, with higher scores indicating greater disability The HAQ-DI is a discrete scale with step values of 0.125, resulting in the HAQ-DI scale containing 25 points. The HAQ-DI has been used in many published RCTs in RA.[23]

2.2 Critique of company's overview of current service provision

The company's overview of current service provision is concise but appropriate and relevant to the decision problem set out in the final NICE scope. The ERG provides a summary of current service provision below.

Clinical guidelines

For people with newly diagnosed RA, NICE CG79[10] 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 cDMARDs" has been used to denote that this involves treatment with multiple cDMARDs simultaneously.

disease-modifying anti-rheumatic drug (tsDMARD). Tofacitinib is available as a 5mg film-coated tablet to be taken by mouth twice a day (BD).

Tofacitinib received a positive opinion from the Committee for Medicinal Products for Human Use (CHMP) on the 26th January 2017 for the treatment of RA. Prior to this approval, the CHMP had adopted a negative opinion for granting marketing authorisation to tofacitinib in 2013 (25th April) which was confirmed on the 22nd July 2013 on the basis of: (1) Serious and unresolved incidence of infection; (2) Uncertainties in the overall safety profile in relation to incidence and severity of infections, malignancies, lymphoma, gastrointestinal perforations, hepatic enzymes elevations/drug-induced liver injury and lipids and cardiovascular risks; (3) Unresolved safety concerns are not offset by the benefits of the treatment. [29] However, the 2017 CHMP opinion concluded that the safety profile of tofacitinib while remaining complex and clinically challenging can now be considered sufficiently characterised for marketing authorisation.

Tofacitinib was added to the EMA's list of medicines under additional monitoring in April 2017.

Laboratory tests are required for patients undergoing treatment with tofacitinib to monitor:

- neutrophils at baseline and after 4 to 8 weeks of treatment and every 3 months thereafter
- lipid parameters after 8 weeks following initiation of therapy
- lymphocytes (at baseline and every 3 months thereafter)
- haemoglobin (at baseline and after 4 to 8 weeks of treatment and every 3 months thereafter).

The ERG's clinical advisors state that these tests would ordinarily be provided in clinical practice for this patient population.

The Summary of Product Characteristics (SmPC)[30] reports the following contraindications for treatment with tofacitinib:

- patients who are allergic or hypersensitive to ingredients of the medicine
- severe hepatic impairment
- pregnant and breast-feeding
- patients with active infections, including localised infections, tuberculosis (TB), serious infections such as sepsis, or opportunistic infection.

The safety and efficacy of tofacitinib in children aged from 2 years to less than 18 years of age have not yet been established.

A number of additional points regarding tofacitinib are emphasised in the SmPC including:

- A higher rate of infections in patients aged 65 and older and diabetic populations.
- A caution that data in the elderly population of 75 years and over are limited.
- A higher rate of herpes zoster (shingles) in Japanese and Korean patients.

Other potentially relevant comparators such as anakinra [KINERET]; baricitinib [OLUMIANT] sarilumab [KEVZARA] and sirukimab, were excluded in the CS as they are either currently unlicensed, unapproved or yet to be assessed by NICE. Baricitinib is currently under assessment by NICE (ID979) for treating moderate-to-severe RA and, like tofacitinib, is an orally administered JAK inhibitor (4mg once per day).

3.4 Outcomes

The outcome measures in the final scope issued by NICE and those considered in the CS are outlined in Table 3.

Table 2: Outcome measures from the NICE scope considered in the CS

Outcomes as per NICE Scope	Outcomes as defined and measured in the CS
Disease activity	Disease Activity Score (DAS28) American College of Rheumatology (ACR)20; ACR50; ACR70 European League Against Rheumatism (EULAR) response
Physical function	Health Assessment Questionnaire-Disability Index (HAQ-DI)
Joint damage, pain	Visual analogue scale (VAS): Patient's assessment of arthritis pain (PAAP)
Mortality	Death within 30 days of last dose of study drug in pooled safety analysis
Fatigue	Functional Assessment of Chronic Illness Therapy-Fatigue (FACIT-F) scale
Radiological progression	Sharp-van der Heidje scale modified Total Score (mTSS)
Extra-articular manifestations of disease	Not provided
Adverse effects of treatment	Pooled incidence rates of 19 trials' intervention arms without comparator
Health-related quality of life.	EuroQol 5-dimension questionnaire (EQ-5D)

3.5 Other relevant factors

Adherence

Adherence to treatment is not measured in the CS however, some potential benefits towards adherence are alluded to. The company states (see CS, page 49) that the mode of administration may be important in adherence to RA treatment and that patients with RA have reported a preference for oral administration over other routes including subcutaneous injection. [31] Whilst the CS references a study[32] which reported that RA patients prefer the oral route of administration to other routes, patient preference

does not necessarily equate with increased adherence. Clinical advice to the ERG was that whilst it may be easier for patients to take oral medication, self-administration in itself may be a contributing factor towards non-adherence, whereas the involvement of a third person can sometimes aid adherence. The CS states some valid potential patient groups where an oral therapy presents a useful alternative to clinicians such as those with impaired hand function who may have problems with self-injection.

Ongoing trials of tofacitinib in RA

Ongoing primary research identified from searching clinicaltrials.gov and relevant to the decision problem is documented in Table 4. Nine ongoing studies were noted to be relevant to the long-term safety and efficacy of tofacitinib and plan completion between April2016 and December 2021.

Table 3: Ongoing trials relevant for tofacitinib in RA

Trial no. Sponsor	Aim	Planned enrolment	Planned completion	Comment on relevance to the decision problem.
NCT02157012 Shinshu University	Phase 4 single arm study to examine the safety and effectiveness after tofacitinib treatment in RA patients	100	April 2016	Recruited exclusively at Japanese sites.
NCT03073109 Pfizer	Study of patient-reported outcomes in RA patients treated with tofacitinib or bDMARDs	320	Mar 2018	Exclusively in Latin American patients.
NCT00413699 Pfizer	Phase 3 study of long-term effectiveness and safety of tofacitinib in RA subjects after participating in another "qualifying" study of tofacitinib (ORAL Sequel)	4500	Dec 2018	Included in the list of non-randomised patients evidence supplied by Pfizer.
NCT02831855 Pfizer	Phase 4 study of methotrexate withdrawal on tofacitinib modified release formulation (11mg QD) versus tofacitinib (11mg QD) plus continued methotrexate treatment	580	Mar 2019	Non-licensed formulation in Europe.
NCT03016884 HaEmek Medical Center, Israel	Phase 4 study evaluating the safety, tolerability, and immunogenicity of Zostavax vaccine in the RA population prior to initiation of biologic/tofacitinib therapy for RA	250	May 2019	Recruited exclusively at Israeli sites.
NCT02092467 Pfizer	Phase 3b/4 post-marketing safety study of tofacitinib compared with ADA and ETN for major cardiovascular adverse events, malignancies, hepatic events, infections, and efficacy parameters.	4400	Aug 2019	
NCT02984020 Pfizer	Korean post-marketing surveillance study for the safety and efficacy of Xeljanz during the post-marketing period as required by the Korean Ministry Of Food And Drug Safety.	3000	Jan 2020	Recruited exclusively at Korean sites.
NCT01932372 Pfizer	Special investigation of tofacitinib 5mg in clinical practice of occurrence of adverse reactions/ factors that may potentially affect safety and efficacy and long-term safety vs other bDMARDs	6000	Mar 2021	Registry study not yet recruiting.
NCT03011281 Hanyang University	Prospective study to evaluate the effectiveness and safety of tofacitinib in clinical practice in Korean RA patients	378	Dec 2021	Exclusively in Korean RA patients.

Source: Clinicaltrials.gov

Simple Disease Activity Index (SDAI), DAS 28-4(ESR), and HAQ-DI over time. This trial was completed in December 2016 and results have recently been published.41 In this trial, tofacitinib plus MTX was found to be non-inferior to adalimumab plus MTX. However, tofacitinib monotherapy failed to demonstrate non-inferiority against tofacitinib plus MTX and adalimumab plus MTX for the primary endpoint of ACR50 response rate. The ERG requested effectiveness data for the ORAL Strategy trial and an updated NMA considering these data (see clarification response,34 question A3). The company's clarification response provided DAS28(ESR) EULAR response data for the full trial population but stated [redacted].

As the results have been published in a peer reviewed publication the ERG note that ORAL Strategy can no longer be considered an ongoing trial and consider that further relevant data from this trial were relevant to the decision problem.

The CS also reported a study "A3921041" (NCT00661661) which was completed in December 2013. This was an open-label, long-term extension study to assess safety, but only included Japanese patients. Clinical advice received by the ERG states that data there may be differences between UK and Japanese clinical populations in terms of tolerance and dosage of cDMARD treatment therefore the ERG considers that data from this trial may not be fully applicable to the decision problem.

4.1.3 Critique of data extraction

The CS reported that data were extracted from eligible publications into a predefined table by 'a reviewer', which is not considered as best practice in undertaking systematic reviews. In response to a request for clarification by the ERG (see clarification response,34 question A4), the company responded that two independent reviewers "were involved" in data extraction and quality assessment.

Data extracted from the four included tofacitinib RCTs reported in the CS, and reported below, were checked by the ERG against published trial papers, and were found to be accurate.

4.1.4 Quality assessment

Quality assessment of the four included tofacitinib RCTs is presented in Section 4.6 and Appendix 4 of the CS. The items assessed were taken from the NICE Single Technology Appraisal: User guide for company evidence submission template.42 These are appropriate criteria for assessing the risk of bias in RCTs. Table 6 presents the company's quality assessment of the tofacitinib trials. It is considered good practice for two reviewers either to independently perform quality assessment or to check assessed items, but this was not reported in the CS. The ERG checked the company's quality assessment against the publications of the RCTs relevant to the decision problem, ORAL Standard

Table 4: Characteristics of included tofacitinib RCTs (adapted from Table 12 of the CS)

Trial acronym and trial number	Population	Intervention, N randomised	Comparators, N randomised	Primary outcome(s)
ORAL Standard NCT00853385	cDMARD experienced and MTX-IR adult patients with active moderate-to-severe RA	Tofacitinib 5mg, oral, BID (with background MTX), N=204	Adalimumab 40mg, SC injection, Q2W (with background MTX), N=204 Placebo to tofacitinib 5mg, oral, BID (with background MTX)† N=56	ACR20 response rate at Month 6 (NRI) HAQ-DI score at Month 3 DAS28(ESR) <2.6 at Month 6 (NRI) (Table 21 of CS)
ORAL Scan NCT00847613	cDMARD experienced and MTX-IR adult patients with active moderate-to-severe RA who are	Tofacitinib 5mg, oral, BID (with background MTX), N=321	Placebo to tofacitinib 5mg, oral, BID (with background MTX)†, N=81	ACR20 response rate at Month 6 (NRI) mTSS score at Month 6 (LE) HAQ-DI score at Month 3 DAS28(ESR) <2.6 at Month 6 (NRI) (Table 27 of CS)
ORAL Sync NCT00856544	DMARD-IR (cDMARD including MTX or bDMARD) adult patients with active moderate-to-severe RA	Tofacitinib 5mg, oral, BID (with background cDMARD), N=315	Placebo to tofacitinib 5mg, oral, BID (with background cDMARDs)†, N=79	ACR20 response rate at Month 6 (NRI) HAQ-DI score at Month 3 DAS28(ESR) <2.6 at Month 6 (NRI) (Table 34 of CS)
ORAL Solo NCT00814307	DMARD-IR (cDMARD including MTX or bDMARD) adult patients with active moderate-to-severe RA	Tofacitinib 5mg, oral, BID, N=243	Placebo to tofacitinib 5mg, oral, BID‡, N=61	ACR20 response rate at Month 3 (NRI) HAQ-DI score at Month 3 (Table 40 of CS)

Abbreviations: ACR = American College of Rheumatology; ADA = adalimumab; bDMARD = biologic disease-modifying anti-rheumatic drug; BID = twice daily; cDMARD = conventional disease-modifying anti-rheumatic drug; IR = inadequate response; LE = linear extrapolation; mTSS = van der Heijde modified total sharp score; MTX = methotrexate; NRI = non-responder imputation; ORAL = Oral Rheumatoid Arthritis Phase 3 Trials; Q2W = twice weekly; RA = rheumatoid arthritis; SC = subcutaneous; TNFi = tumour necrosis factor inhibitor; TOF = tofacitinib.

Footnote: †Patients receiving placebo advanced to TOF 5 mg at Month 3 if trial response criteria were not met (defined as 20% reduction in number of tender and swollen joints) or Month 6 regardless of response. ‡All patients receiving placebo advanced to a TOF 5 mg at Month 3

In all four placebo-controlled trials, data for the placebo comparator group is presented in the CS as a "combined placebo group" because patients crossed over to receive either 5 mg (licensed dose) or 10 mg of tofacitinib but results are not provided for the licenced 5 mg dose separately. An early escape design allowed that, at Month 3, placebo non-responders advanced to either 5 mg or 10 mg tofacitinib and at Month 6, all patients receiving placebo advanced to either 5mg or 10mg tofacitinib "*in order to minimise the time patients spent on ineffective treatment*" (CS, page 89). Additionally in the ORAL Standard, Scan and Sync trials an "advancement penalty" was applied whereby patients who did not meet the response criteria at Month 3 were considered to be non-responders for the remainder of the trial. This non-responder imputation (NRI) was also applied to the analysis of patients deemed to be non-responders in the tofacitinib treatment groups at Month 3.

Table 5: Eligibility criteria for the tofacitinib RCTs (reproduced from Table 14 of the CS)

Trial acronym and trial number	ORAL Standard NCT00853385	ORAL Scan NCT00847613	ORAL Sync NCT00856544	ORAL Solo NCT00814307
Inclusion criteria	• Adults aged ≥18 years with a diagnosis of active RA†, consistent with the ACR 1987 Revised Criteria[15] • Ongoing treatment with MTX for ≥4 months with stable dosing (7.5–25 mg/week) ≥6 weeks before receiving the study drug; doses <15 mg were allowed in the case of intolerance or toxicity from higher doses • An inadequate response to MTX (defined as sufficient residual disease activity to meet entry criteria)	• Adults aged ≥18 years with a diagnosis of active RA†, consistent with the ACR 1987 Revised Criteria[15] • Ongoing treatment with MTX for ≥4 months with stable dosing (7.5–25 mg/week) ≥6 weeks before receiving the study drug; doses <15 mg were allowed in the case of intolerance or toxicity from higher doses • An inadequate response to MTX (defined as sufficient residual disease activity to meet entry criteria) • Evidence of ≥3 distinct joint erosions on posteroanterior hand and wrist radiographs or anteroposterior foot radiographs as determined by the investigator, or, if radiographic evidence of joint erosions was unavailable, IgM RF+ or antibodies to CCP	• Adults aged ≥18 years with a diagnosis of active RA‡, consistent with the ACR 1987 Revised Criteria[15] • Ongoing treatment with ≥1 cDMARD therapy – patients receiving MTX required ≥4 months of treatment, with stable dosing (≤25 mg/week) ≥6 weeks before receiving the study drug • An inadequate response to ≥1 cDMARD or bDMARD ([redacted]	• Adults aged ≥18 years and had received a diagnosis of active RA†, consistent with the ACR 1987 Revised Criteria[15] • Discontinued all DMARDs except stable doses of anti-malarial agents • An inadequate response to ≥1 cDMARD or bDMARD (lack of efficacy or occurrence of toxicity)

Exclusion criteria	• Haemoglobin <9.0 gm/dL • Haematocrit <30% • White blood cell count $<3.0x10^9$/L • Absolute neutrophil count $<1.2x10^9$/L • Platelet count $<100x10^9$/L • eGFR rate ≤40 ml/min • AST or ALT levels >1.5 x Upper limit of normal • A history of another autoimmune rheumatic disease except Sjögren's syndrome • Infection that required hospitalisation or parenteral antimicrobial therapy within 6 months of randomisation • Infection requiring antimicrobial therapy within 2 weeks of randomisation • Recurrent or disseminated herpes zoster infection • Recent, current, or chronic infection, including HBV, HCV or HIV • Current infection or evidence of active or inadequately treated infection with Mycobacterium tuberculosis • History of lymphoproliferative disorder or malignancy except for adequately treated non-metastatic basal/squamous cell cancer of the skin or cervical carcinoma in situ • Prior treatment with lymphocyte-depleting therapies or alkylating agents ORAL Standard only: • Prior treatment with ADA • Lack of response to prior anti-TNF biologic treatment • Current treatment with other anti-rheumatic agents, including biologic agents

Abbreviations: ACR = American College of Rheumatology; ADA = adalimumab; ALT = alanine aminotransferase; AST = aspartate aminotransferase; bDMARD = biologic disease-modifying anti-rheumatic drug; CCP = cyclic citrullinated peptide; cDMARD = conventional disease-modifying anti-rheumatic drug; CRP = C-reactive protein; eGFR = estimated glomerular filtration rate; ESR = erythrocyte sedimentation rate; HBV = hepatitis B virus; HCV = hepatitis C virus; HIV = human immunodeficiency virus; MTX = methotrexate; ORAL = Oral Rheumatoid Arthritis Phase 3 Trials; RA = rheumatoid arthritis; RF = rheumatoid factor; TNF = tumour necrosis factor.

Footnote: †Active disease was defined as the presence of ≥6 tender or painful joints (of 68 joints examined) and ≥6 swollen joints (of 66 joints examined) and either an ESR ≥28 mm/hr (Westergren method) or a CRP level >7 mg/L. ‡Active disease was defined as the presence of ≥4 tender or painful joints (68 joints examined) and ≥4 swollen joints (of 66 joints examined) and either an ESR ≥28 mm/hr or a CRP level >66.7 nmol/L

Table 6: **Baseline characteristics of participants of ORAL Sync (adapted from Table 18 of the CS)**

ORAL Sync		Placebo to TOF 5mg (N=79)	Placebo to TOF 10mg (N=80)	TOF 5 mg (N=315)
Gender, n (%)	Female, n (%)	[redacted] (79.7)	[redacted] (75.0)	[redacted] (83.8)
	Male, n (%)	[redacted] (79.7)	[redacted] (25.0)	[redacted] (16.2)
Race, n (%)	White	[redacted] (60.8)	[redacted] (55.0)	[redacted] (54.9)
	[redacted]	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]	[redacted]
Region of origin, %	Europe	31.7	28.8	28.9
	North America	22.8	18.8	16
	Latin America	13.9	13.8	14.2
	Rest of world	31.7	38.8	40.9
Age, years (SD)		50.8 (11.2)	53.3 (10.8)	52.7 (11.7)
Mean duration of RA	Years (range)	9.5 (0.3–39.3)	10.2 (0.3–49.0)	8.1 (0.2–39.9)
Rheumatoid factor	n	[redacted]	[redacted]	[redacted]
	Positive, n (%)	[redacted] (73.1)	[redacted] (72.2)	[redacted] (73.9)
Anti-CCP	n	[redacted]	[redacted]	[redacted]
	Positive, n (%)	[redacted]	[redacted]	[redacted]
Tender and swollen joints	n	[redacted]	[redacted]	[redacted]
	Tender joints, mean (SD)	27.2 (16.8)	21.9 (13.0)	25.0 (15.3)
	Swollen joints, mean (SD)	14.6 (9.7)	13.9 (8.6)	14.5 (10.3)
DAS28(ESR)	n	[redacted]	[redacted]	[redacted]
	Mean (SD)	6.44 [redacted])	6.14 [redacted]	6.27 [redacted]
DAS28-3(CRP)	n	[redacted]	[redacted]	[redacted]
	Mean (SD)	[redacted]	[redacted]	[redacted]
HAQ-DI score	n	[redacted]	[redacted]	[redacted]
	Mean (SD)	1.45 (0.64)	1.24 (0.66)	1.44 (0.69)
Prior therapy	TNF inhibitor, n (%)	[redacted] (6.3)	[redacted] (6.3)	[redacted] (7.3)
	Non-TNF inhibitor bDMARD, n (%)	[redacted](7.6)	0	[redacted](2.2)
	MTX, n (%)	[redacted] (83.5)	[redacted] (82.5)	[redacted] (86.7)
	Non-MTX cDMARD, %	55 (69.6)	62 (77.5)	232 (73.7)
	Failed DMARDs, mean	1.3	1.4	1.4
Concomitant therapy = n (%)	MTX	61(77.2)	64 (80.0)	250 (79.4)
	1 cDMARD	[redacted] (73.4)	[redacted] (62.5)	[redacted] (66.7)
	≥2 cDMARDs	[redacted] (25.3)	[redacted] (37.5)	[redacted] (33.3)
	NSAIDs	[redacted] (72.2)	[redacted] (63.8)	[redacted] (75.9)
	Systemic CCS	[redacted] (59.5)	[redacted] (58.8)	[redacted] (61.9)
	Lipid-lowering medication	[redacted]	[redacted]	[redacted]

Abbreviations: bDMARD = biologic disease-modifying anti-rheumatic drug; CCP = cyclic citrullinated peptide; CCS = corticosteroid; cDMARD = conventional disease-modifying anti-rheumatic drug; CRP = C-reactive protein; DAS28 = Disease Activity Score in 28 joints; ESR = erythrocyte sedimentation rate; HAQ-DI = Health Assessment Questionnaire-disability index; MTX = methotrexate; NSAID = non-steroidal anti-inflammatory drug; ORAL = Oral Rheumatoid Arthritis Phase 3 Trials; RA = rheumatoid arthritis; SD = standard deviation; TNF = tumour necrosis factor; TOF = tofacitinib.
Footnote: [†]In the ORAL trial programme Asian refers to Japanese and Korean patients.

All four RCTs employed modified intention-to-treat (mITT) analyses for effectiveness measures, comprising all randomised patients who received at least one dose of the study drug. All randomised patients in ORAL Standard (n=717) were included in the mITT analyses. Within ORAL Scan, 797/800 (99.6%) patients were included in the mITT analyses. Within ORAL Sync, 792/795 (99.6%) patients were included in the mITT analyses. Within ORAL Solo, 610/611 (99.8%) patients were included in the mITT analyses. All four RCTs are analysed with non-responder imputation and missing data are accounted for using last observation carried forward (LOCF) (see CS, pages 154-159).

4.2.2 Efficacy results for tofacitinib

ACR response data

ACR20 response data for the four included tofacitinib RCTs (ORAL Standard, Scan, Sync and Solo) are reported in Table 15, Table 16, Table 17 and Table 18 respectively. A co-primary outcome for ORAL Standard, ORAL Scan and ORAL Sync was the proportion of patients achieving an ACR20 response at six months. A co-primary outcome for ORAL Solo was the proportion of patients achieving an ACR20 response at three months. For ACR20, all four RCTs found a statistically significant advantage for tofacitinib 5mg BID compared with the combined placebo group: ORAL Standard, 51.5% vs 28.3% ($p<0.001$); ORAL Scan, 51.5% vs 25.3% ($p<0.001$); ORAL Sync 52.7% vs 31.2% ($p<0.001$); ORAL Solo 59.8% vs 26.7% ($p<0.001$) (see Table 15, Table 16, Table 17, and Table 18).

ACR50 responses for tofacitinib versus placebo were ORAL Standard, [redacted]% vs [redacted]% (p[redacted]); ORAL Scan, 32.4% vs 8.4% ($p<0.001$); ORAL Sync, [redacted]% vs [redacted]% ($p\leq$[redacted]); ORAL Solo, 31.1% vs 12.5% ($p<0.001$) (data taken from the CS, Tables 23, 29, 36 and 41).

ACR70 responses for tofacitinib versus placebo were ORAL Standard, [redacted]% vs [redacted]% (p[redacted]); ORAL Scan, 14.6% vs 1.3% ($p<0.001$); ORAL Sync, [redacted]% vs [redacted]% (p[redacted]); ORAL Solo, 15.4% vs 5.8% ($p<0.001$) (data taken from the CS, Tables 23, 29, 36 and 41).

For ORAL Standard, the CS (page 106) reported that in terms of comparison between tofacitinib and adalimumab:

“[redacted]”

For the recently completed, head-to-head trial, ORAL Strategy, the CS (page 251) reported the preliminary primary endpoint data (ACR50 response) for tofacitinib plus MTX vs adalimumab plus MTX vs tofacitinib monotherapy. Table 14 shows that tofacitinib plus MTX, but not tofacitinib monotherapy, was non-inferior to adalimumab plus MTX. Data were provided in the CS as academic in confidence but have subsequently been published in an open access peer reviewed publication.[41]

Table 7: ORAL Strategy ACR50 response rates at Month 6 including non-inferiority results (adapted from Table 89 of the CS)

Outcome		**TOF 5 mg Monotherapy (N=384)**	**TOF 5 mg + MTX (N=376)**	**ADA 40 mg + MTX (N=386)**
ACR50 response rate at Month 6, n (%)		147 (38.28)	173 (46.01)	169 (43.78)
Differences in ACR50 response rate				
Comparing with ADA 40 mg + MTX	Absolute difference (TOF – ADA), %	-5.50	2.23	-
	98.34% CI*	-13.98, 2.98	-6.40, 10.86	-
	Non-inferiority criteria met?	No	Yes	-
	p-value†	0.0512	<0.0001	-
Comparing with TOF 5 mg + MTX	Absolute difference (TOF mono – TOF+MTX), %	-7.73	-	-
	98.34% CI*	-16.29, 0.83	-	-
	Non-inferiority criteria met?	No	-	-
	p-value†	0.2101	-	-

Abbreviations: ACR, American College of Rheumatology; ADA, adalimumab; CI, confidence interval; MTX, methotrexate; TOF, tofacitinib.
†p-values are from non-inferiority hypothesis testing. The p-values are multiplicity-adjusted and should be compared with 0.05.
* Non-inferiority between groups was shown if the lower bound of the 98·34% CI of the difference between comparators was larger than –13·0%

In the corresponding journal publication (Fleischman *et al.*, 2017)[41] the authors claim that the results suggest that in patients with an inadequate response to MTX, the addition of tofacitinib or adalimumab is equally efficacious and more likely to be effective than switching to tofacitinib monotherapy. The paper further asserts, "[t]*he present analysis suggests that adding tofacitinib 5 mg BID to MTX is as effective as adding adalimumab, a TNFi, to MTX*". The ERG notes that non-inferiority trials do not provide evidence that interventions are therapeutically equal, which is instead the purpose of an equivalence trial. Non-inferiority trials aim to determine whether one treatment is not statistically worse than another. In this case, non-inferiority was only demonstrated for tofacitinib

combination therapy but tofacitinib monotherapy was not found to be non-inferior in the relevant patient population for the current decision problem.

EULAR response data

The CS estimated EULAR response criteria from DAS28 scores as a good or moderate EULAR response (described in the CS as an improvement in DAS28 from baseline) for ORAL Standard, ORAL Scan and ORAL Sync at six months and for ORAL Solo at three months. For this outcome, the responses for tofacitinib 5mg BID compared with the combined placebo group were ORAL Standard, [redacted]% vs [redacted]% (*p*[redacted]); ORAL Scan, vs [redacted]% [redacted]% (*p*[redacted]); ORAL Sync vs [redacted]% [redacted]% (*p*[redacted]); ORAL Solo [redacted]% vs [redacted]% (*p*[redacted]) (see CS, Tables 25, 31, 38 and 43).

Change from baseline in HAQ-DI scores

Mean change from baseline in HAQ-DI scores for the four included tofacitinib RCTs are shown in Table 15, Table 16, Table 17 and Table 18. The primary outcome for ORAL Standard, ORAL Scan, ORAL Sync and ORAL Solo was the mean change from baseline in HAQ-DI score at three months. For this outcome, ORAL Standard, ORAL Sync and ORAL Solo found a statistically significant advantage for tofacitinib 5mg BID compared with the combined placebo group: ORAL Standard, –0.55 vs –0.24 ($p<0.001$); ORAL Sync–0.46 vs –0.21 ($p<0.001$); ORAL Solo–0.50 vs–0.19 ($p<0.001$) (CS Tables 21, 27, 34 and 40). For ORAL Scan, the HAQ-DI scores for tofacitinib 5 mg BD versus placebo were not statistically significant (*p*-value not declared).

Mean change from baseline in HAQ-DI scores to six months for tofacitinib 5mg BID compared with the combined placebo group were: ORAL Standard, [redacted] vs [redacted] (*p*[redacted]); ORAL Scan, [redacted] vs [redacted] (*p*[redacted]); ORAL Sync, [redacted] vs [redacted] (*p*[redacted]); ORAL Solo,–0.50 vs –0.19 ($p<0.001$) (3 month data only available for ORAL Solo) (see CS, Tables 24, 30, 37 and 40).

DAS28(ESR) <2.6 and ≤3.2 response

DAS28(ESR) <2.6 response data for the four included tofacitinib RCTs are shown in Table 15, Table 16, Table 17 and Table 18. The primary outcome for ORAL Standard, ORAL Scan and ORAL Sync was the proportion of patients achieving a DAS28(ESR) <2.6 response at six months. The primary outcome for ORAL Solo was the proportion of patients achieving a DAS28(ESR) <2.6 response at three months. The proportions of patients achieving a response for tofacitinib 5mg BID compared with the combined placebo group were: ORAL Standard, 6.2% vs 1.1% (*p*[redacted]); ORAL Scan, 7.2% vs 1.6% (statistical significance was not declared); ORAL Sync 9.1% vs 2.7% ($p=0.0038$); ORAL Solo 5.6% vs 4.4% ($p=0.62$) (CS Tables 25, 31, 34 and 40).

The proportions of patients achieving a DAS28(ESR) ≤3.2 response for tofacitinib 5mg BID compared with the combined placebo group were: ORAL Standard, [redacted]% vs [redacted]% (*p*[redacted]); ORAL Scan, [redacted]% vs [redacted]% (*p*[redacted]); ORAL Sync [redacted]% vs [redacted]% (*p*[redacted]); ORAL Solo 12.5% vs 5.3% (*p*<0.001) (see CS, Tables 25, 27, 38 and 43).

Table 8: Summary of primary efficacy results for ORAL Standard (adapted from CS Table 21)

Outcome		**Placebo to tofacitinib 5mg or 10mg BID**	**Tofacitinib 5mg BID**	**Adalimumab 40mg SC Q2W**
ACR20 response rate at Month 6 (NRI with advancement penalty)	n	106	196	199
	Response rate, n (%)	30 (28.3)	101 (51.5)	94 (47.2)
	Difference from placebo, %	-	[redacted]	[redacted]
	95% CI for difference	-	[redacted]	[redacted]
	p-value†	-	<0.001	<0.001
HAQ-DI score at Month 3	n	98	188	190
	LS mean change from baseline	–0.24	–0.55	–0.49
	LS mean difference from placebo	-	[redacted]	[redacted]
	95% CI for difference	-	[redacted]	[redacted]
	p-value†	-	<0.001	<0.001
DAS28(ESR) <2.6 at Month 6 (NRI with advancement penalty)	n	92	177	178
	Response rate, n (%)	1 (1.1)	11 (6.2)	12 (6.7)
	Difference from placebo, %	-	[redacted]	[redacted]
	95% CI for difference	-	[redacted]	[redacted]
	p-value†	-	[redacted]	[redacted]

Abbreviations: ACR = American College of Rheumatology; ADA = adalimumab; CI = confidence interval; DAS28 = Disease Activity Score in 28 joints; ESR = erythrocyte sedimentation rate; FAS = full analysis set; HAQ-DI = Health Assessment Questionnaire-disability index; LS = least squares; NRI = non-responder imputation; Q2W = twice weekly; SC = subcutaneous; TOF = tofacitinib.

Footnote: †p-value is subject to the step-down approach

Table 15 shows that both tofacitinib and adalimumab were significantly superior to placebo for the ACR20 and DAS28(ESR) outcomes at 6 months and HAQ-DI at 3 months.

Table 9: Summary of primary efficacy results for ORAL Scan (adapted from CS Table 27)

Outcome		Placebo to tofacitinib 5mg or 10mg BID	Tofacitinib 5mg BID
ACR20 response rate at Month 6 (NRI with advancement penalty)	n	[redacted]	[redacted]
	Response rate, n (%)	[redacted] (25.3)	[redacted] (51.5)
	Difference from placebo, %	-	[redacted]
	95% CI for difference	-	[redacted]
	p-value[†]	-	<0.001
HAQ-DI score at Month 3	n	[redacted]	[redacted]
	LS mean change from baseline	–0.15	–0.40
	LS mean difference from placebo	-	[redacted]
	95% CI for difference	-	[redacted]
	p-value[†]	-	Not declared[‡]
DAS28(ESR) <2.6 at Month 6 (NRI with advancement penalty)	n	[redacted]	[redacted]
	Response rate, n (%)	[redacted] (1.6)	[redacted] (7.2)
	Difference from placebo, %	-	[redacted]
	95% CI for difference	-	[redacted]
	p-value[†]	-	Not declared[‡]
mTSS score at Month 6 (LE)	n	[redacted]	[redacted]
	LS mean change from baseline	0.47	0.12
	LS mean difference from placebo	-	[redacted]
	95% CI for difference	-	[redacted]
	p-value[†]	-	0.0792

Abbreviations: ACR = American College of Rheumatology; ADA = adalimumab; CI = confidence interval; DAS28 = Disease Activity Score in 28 joints; FAS = full analysis set; HAQ-DI = Health Assessment Questionnaire-disability index; LE = linear extrapolation; LS = least squares; NRI = non-responder imputation; mTSS = van der Heijde modified total sharp score; TOF = tofacitinib.
Footnote:[†]p-value is subject to the step-down approach. [‡]Due to the step-down procedure applied to primary efficacy outcomes, significance was not declared for the HAQ-DI score or DAS28(ESR) <2.6 for TOF 5 mg. Nominal p-values (TOF 5 mg vs placebo) for these outcomes were <0.001 and 0.0034, respectively

Table 15 shows that both tofacitinib and adalimumab were significantly superior to placebo for the ACR20 and DAS28(ESR) outcomes at 6 months and HAQ-DI at 3 months.

Table 16 shows that ACR20 was the only outcome where tofacitinib 5 mg BD was declared to be significantly superior to placebo. A step-down approach was used for statistical testing in the order of ACR20, mTSS, HAQ-DI and then DAS28-4(ESR) <2.6. As the mean change from baseline in mTSS score at Month 6 was not significantly different between the tofacitinib 5 mg group (0.12) and the combined placebo group (0.47; p=0.0792), no statements regarding statistical significance could be declared for HAQ-DI score or DAS28-4(ESR) <2.6 for tofacitinib 5 mg.

Table 10: ORAL trials January 2016 data set analysis: tofacitinib safety data (replicated from clarification response, question A1[34])

Event Term	Total number of events	Number of patients affected	Incidence per 100 patient exposure years
Serious Infection Events	[redacted]	[redacted]	[redacted]
Drug Induced Liver Injury (Cases meeting Hy's law†)	[redacted]	[redacted]	[redacted]
Gastrointestinal Perforation Events	[redacted]	[redacted]	[redacted]
Treatment discontinuations as a result of an Adverse Event	[redacted]	[redacted]	[redacted]
All-cause mortality	[redacted]	[redacted]	[redacted]
Herpes Zoster infection	[redacted]	[redacted]	[redacted]
Interstitial Lung Disease	[redacted]	[redacted]	[redacted]
Malignancies			
All Cancers (other than non-melanomatous cancers of the skin)	[redacted]	[redacted]	[redacted]
Lymphoma	[redacted]	[redacted]	[redacted]
Non-melanomatous cancers of the skin	[redacted]	[redacted]	[redacted]
Breast Cancer (Female patients only)	[redacted]	[redacted]	[redacted]
Lung Cancer	[redacted]	[redacted]	[redacted]
Melanoma	[redacted]	[redacted]	[redacted]

Footnote: †prognostic indicator that a pure drug-induced liver injury (DILI) leading to jaundice, without a hepatic transplant, has a case fatality rate of 10% to 50%.

According to the data presented in the company response, the most commonly recorded AE was herpes zoster infection, with an estimated incidence rate per 100 patient years of [redacted]) (Table 23). However, the ERG's own search for AEs in Medline retrieved a study by Winthrop et al., (2014) who reviewed the tofacitinib RA development programme from the Phase II, III and long-term extension studies. This earlier data cut of March 2011 reported the incidence rate of herpes zoster was 4.3 per 100 patient years but was substantially higher within Asia (7.7 per 100 patient years). Clinical advice received by the ERG suggested that increased risk of herpes zoster is elevated about 2-fold in RA generally and the experts considered an increased risk by treatment as therefore more worrying as some instances can be serious, particularly in the elderly. Neither the CS nor the company's response to the clarification letter not provides incidence rates for the comparators arms, instead an analysis is presented which shows that the rate of herpes zoster is relatively stable over time (measured at 6-monthly intervals

tofacitinib 5 mg whilst the proportion of patients experiencing ≥1 treatment-related AE at 3 months in the ORAL Standard, Scan and Sync (tofacitinib plus methotrexate) trials was [redacted] respectively (see CS, Appendix 2). The ERG has tabulated selected AE data deemed as related to the study drug for the tofacitinib treatment arms (data from both 5 mg and 10 mg arms) for the four key ORAL trials. As can be observed in Table 27, the three-tofacitinib combination trials have higher incidences of the selected treatment-related AEs than the monotherapy trial (ORAL Solo).

Table 11: Tofacitinib-related adverse event (data extracted from Appendix 2 of the CS)

Number experiencing event/ Number of patients in tofacitinib (5 mg and 10 mg) treatment arms				
	ORAL Standard	ORAL Scan	ORAL Sync	ORAL Solo
Treatment related SAEs between 0-6 months	[redacted]	[redacted]	[redacted]	[redacted]
Discontinuation due to AEs between 0-6 months	40/405 (9.9%)	53/637 (8.3%)	40/633 (6.3%)	14/488 (2.9%)
Deaths attributed to study treatment	1	5	3	0

Interestingly the recently published journal paper for the ORAL Strategy trial[41] describes this same issue (which is not drawn in the CS) when the authors state that "*concomitant csDMARDS augment the risk of herpes zoster with tofacitinib.*" They cite an abstract from a study funded by Pfizer which found that "*concomitant use of nonbiologic DMARDs or GCs appears to increase the risk and overall IR per 100* [patient years] *of HZ from 0.56 to 4.82 with 5 mg BID*".[50] This study, published in 2015, is not referenced in the CS.

The ERG considers that a higher toxicity profile of tofacitinib plus methotrexate cannot be fully characterised in a pooled analysis with associated incidence rates from both dosing regimens, as combining the monotherapy and combination therapy trials potentially dilutes the apparent incidence of treatment-related adverse events that occur in tofacitinib combination therapy.

4.3 Critique of trials identified and included in the network-meta-analysis

4.3.1 Included trials for the network meta-analysis

NMAs were performed separately for the cDMARD-IR and bDMARD-IR population. Trials other than the tofacitinib RCTs (ORAL Standard, ORAL Scan, ORAL Sync, ORAL Solo and ORAL Step)

that were included in the NMA are listed in Table 28 (cDMARD-IR population) and Table 29 (bDMARD-IR population) below.

Quality assessments of the included trials (other than ORAL Standard, ORAL Scan, ORAL Sync, ORAL Solo and ORAL Step) were presented in Appendix 4 of the CS. Appropriate quality assessment items were used, however, it was unclear for the double-blind trials in Appendix 4 of the CS, who exactly was blinded (i.e., patients, physicians, outcome assessors). In response to a request for clarification from the ERG regarding who was blinded in the double-blind trials (see clarification response,[34] question A6), the company stated:

> *"patients and investigators were blind in six trials (ADACTA[51], AUGUST II[52], LITHE[53, 54], OPTION[55], PLANETRA [56], Van de Putte 2004[57]); patients and outcome assessors were blind in four trials (DE019, RAPID 1, RAPID 2, GO-FORTH); patients, care providers, and investigators were blind in one trial (GO-FORWARD); patients, care providers, investigators, and outcome assessors were blind in 11 trials (ACT-RAY, ATTEST, CERTAIN, Choe 2015, Emery 2015, Fleischmann 2012, GO-FURTHER, HERA, J-RAPID, Kremer 2012, Li 2015, SATORI); and patients, investigators, and other study personnel, except for pharmacists were blind in one trial (START)."*

It was not reported who was blinded in three of the "double-blind" trials (CHANGE[58], Kim 2007[59] and Van de Putte *2004*[57]).

Trials in the analysis of the cDMARD-IR population were largely the same as those in the NMA undertaken by the independent Assessment Group (AG) in TA375. However, there were some exceptions, which have been grouped into the following categories: (i) trials in the CS that were not included in TA375, and; (ii) trials included in TA375 but excluded from the CS. A similar comparison could not be made for the bDMARD-IR population, as this was not the focus of TA375.

Trials included the CS not in TA375 NMA

In total, 10 trials were included the CS that were not included in the base case analysis of TA375. HERA[61] was published after the search date for TA375. Fleishmann 2012,[62] GO-AFTER,[63] Kremer 2012[64] and RADIATE[65], were excluded from TA375 as participants in these trials had received prior biologic therapy. J-RAPID[66] was excluded as separate 6-month data were not reported for those with concomitant cDMARDs and monotherapy. Four trials were only included in TA375 sensitivity analyses as trial participants had received prior biologics (LITHE,[53, 54] OPTION,[67] RAPID 1,[68, 69] RAPID 2[70]).

Trials in TA375 NMA not in the CS base case

The ERG identified 19 trials that had been included in TA375 that were either excluded or not included in the CS. Of these, 12 trials in TA375 were identified as potentially relevant and full texts were scrutinised by the ERG. Possible reasons for exclusion identified by the ERG for all 12 studies are presented in Table 30.

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

NMAs were performed separately for the cDMARD-IR and bDMARD-IR population using a Bayesian approach for EULAR response at Month 6 and change from baseline HAQ-DI score at Month 6. For the continuous outcome, HAQ-DI, an identity-link function model was used in the NMA. For the ordered categorical EULAR response, a binomial likelihood with logit link-function model was used for the cDMARD-IR population by dichotomising the data, and a multinomial likelihood with probit link function model was used for bDMARD-IR population. The CS also explores the probit link function model for the cDMARD-IR population in a scenario analysis. The choice of the link function was based on the performance of convergence of the Markov chain Monte Carlo (MCMC). The choice between the fixed effect and random effects model was based on the deviance information criterion (DIC). Table 31 provides a summary of the model used for each outcome measure in the two populations.

Table 12: The model used for each analysis in the CS

Population	Outcome	Model
cDMARD-IR	EULAR response (moderate)	binomial logit (fixed effect)
	EULAR response (good)	binomial logit (fixed effect)
	EULAR response (at least moderate)	binomial logit (random effects)
	HAQ-DI	identity (random effects)
bDMARD-IR	EULAR response	multinomial probit (fixed effect)
	HAQ-DI	identity (fixed effect)

The ERG disagrees with the approach of using two different models for EULAR response in the two populations based on the performance of the convergence of the MCMC. When data are sparse, poor convergence may be caused by the use of a reference/vague prior. The choice of the likelihood function/link function should be based on the data generating process. A multinomial likelihood with probit link function is preferred to a binomial likelihood with logit link function for the ordered categorical EULAR data because it accounts for natural ordering and correlations between the EULAR categories. This is important to the decision problem when EULAR results are used to populate the economic model. When data are sparse, comparing DIC of a fixed effect model with DIC of a random effects model using a reference/vague prior for the between-study standard deviation may not be appropriate since the reference/vague prior may lead to implausible posterior uncertainty for the results. The choice between the fixed effect and random effects model should be determined by the objective of the analysis and the conduct of the included studies. The fixed effect model was used for a moderate EULAR response and a good response, but the random effects model was used for at least a moderate

response in the cDMARD-IR population. It may not be reasonable to believe that heterogeneity exists in at least a moderate EULAR response network but not in a moderate response or a good response network.

In response to a request for clarification (question A11), the company clarified that placebo + cDMARD/cDMARD was used as the reference treatment across all the NMAs.

For tofacitinib (TOF) trials with early escape, two non-responder imputation (NRI) approaches were applied. Estimate 1 of treatment effect was calculated by applying NRI to Month 3 non-responders from the placebo arm (termed NRI without advancement penalty). Estimate 2 of treatment effect was calculated by applying NRI to Month 3 placebo non-responders as well as the Month 3 TOF non-responders (termed NRI with advancement penalty). The primary analysis for the ORAL Standard, Scan and Sync trials was based on NRI with advancement penalty (Estimate 2).

Estimate 1 was used in the base case NMA for the ORAL Standard, Scan and Sync trials with the justification that, using the data combined from these three trials, [redacted] of non-responders treated with TOF at Month 3 subsequently developed a response to treatment at Month 6. The CS states that clinical expert opinion estimates that less than 10% of the Month 3 placebo-treated non-responders would have subsequently developed a EULAR response by Month 6 (CS page 156). Estimate 1 was also used in the base case NMA for the ORAL Solo and Step trials with the reason that it is expected that few patients would go on to develop any subsequent response to treatment beyond that already seen by Month 3 (CS page 158) in the absence of any form of active DMARD treatment. The ERG believes that Estimate 1 overestimates the relative treatment effect of TOF and Estimate 2 underestimates the treatment effect of TOF.

In response to a request for clarification (question A12), the company stated that there was a typographical error in the CS regarding the prior used for the treatment effect relative to the reference treatment. The vague prior used for the relative treatment effect was a normal distribution with mean 0 and variance 100^2. In RE models, a uniform [0, 5] prior was used for the between-study standard deviation. The ERG notes that when data are sparse, this uniform prior would lead to implausible posterior uncertainty in the results.

The I^2 statistic was used to assess the heterogeneity for the pairwise treatment comparisons. [redacted]

Because a probit model was used in the bDMARD-IR population for EULAR response, it was not clear how the OR was calculated in this case. In response to a request for clarification from the ERG (question A11), the company stated that the WinBUGS code presented included code for generating the absolute treatment effects but these were not generated. Hence, it was still unclear how ORs were calculated from the probit model.

The base case NMA results in the CS should be interpreted with caution since Estimate 1 (NRI without advancement penalty) was used for calculating the relative treatment effect of TOF in the ORAL trials, which overestimated the relative treatment effect of TOF in these trials. A fixed effect model was used for moderate EULAR response, good EULAR response in the cDMARD-IR population and all the outcomes in the bDMARD-IR population, which underestimated treatment uncertainty. Two different models were used for EULAR response in the two populations.

To incorporate etanercept into the cDMARD-IR networks, the company assumed that the intensified cDMARD arm in the LARA study was the same as the cDMARD node, based on the assumptions involved in incorporating LARA to the central node were less of a risk to bias in the network than changing the inclusion criteria for the NMA to include the SWEFOT trial (disease duration <1 year) in the base case analysis. The ERG notes that this may not be an appropriate assumption to make, because this could overestimate the treatment effect of cDMARD.

Six sensitivity analyses were performed in the CS, which included:

1. Exclusion of predominantly Asian populations trials/lower dose MTX
2. Exclusion of trials that included patients with prior bDMARD exposure
3. Exclusion of trials with milder disease
4. Separating intensified cDMARDs from central node
5. Alternative modelling approach (probit) for cDMARD-IR
6. Alternative modelling approach (probit) for cDMARD-IR, using Estimate 2

The company concluded that results were sensitive to the trials included in the base case network, but less influenced by the modelling approach.

The ERG requested the company to perform additional analysis for EULAR response in both populations (clarification question A7) with the following settings:

- Using a random effects probit model with an informative prior for the between-study variance (log normal with mean of -2.56 and variance of 1.74^2, proposed by Turner *et al.*, (2012).[113] The log normal is truncated so that the OR in one study would not be ≥50 times than in another, and re-scaled to match the probit scale).

- EULAR response for ORAL trials derived using DAS ESR with all trial data by applying non-responder imputation Estimate 2 in the CS Table 53. Use the individual EULAR results from trials in the NMA, i.e. not pooling individual patient-level data from ORAL trials.
- Excluding studies which only reported DAS (i.e. did not report EULAR) from the NMA.
- Not assuming intensified DMARD arm is equivalent to the central DMARD node in the LARA trial and including the SWEFOT trial. [redacted]Choosing PBO plus cDMARD/cDMARD as the reference treatment (treatment 1) in the analyses.

The ERG also requested a sensitivity analysis for the requested NMA as above by excluding patients with prior biologic use in the ORAL trials and excluding studies that enrolled a proportion of patients with prior bDMARD use (clarification question A8). In addition to the two analyses the ERG has requested, the company also provided the results using the settings suggested by the ERG as above but applying Estimate 1 (NRI without advancement penalty) to the ORAL trial [redacted] to [redacted] show the EULAR results from the additional analyses conducted by the company (clarification question A7 and A8). All the results were interventions relative to cDMARD on the probit scale, with larger negative numbers being associated with better health outcomes.

Using Estimate 2 (NRI with advancement penalty), which is consistent with the primary analysis of the ORAL Standard, Scan and Sync trials, the effect of TOF plus cDMARD was the smallest among the bDMARDs in the cDMARD-IR population (Figure 2). Using Estimate 1 (NRI without advancement penalty), the effect of TOF + cDMARD compared to cDMARD was smaller than that of TCZ, CTZ, GOL, ETN and ETN's biosimilars in combination with cDMARD, but larger than ADA, ABT, IFX and IFX's biosimilars in combination with cDMARD in the cDMARD-IR population (Figure 2).

For TOF as monotherapy, the effect of TOF compared with cDMARD was the smallest among the active treatments using Estimate 2, but had a larger effect than intensified cDMARD and ETN using Estimate 1 in the cDMARD-IR population (Figure 3).

The analyses including patients with and without prior biologics use provide very similar results for the cDMARD-IR population, except that the treatment effect of TCZ plus cDMARD versus cDMARD reduced noticeably using the studies without prior biologics and the effect of ADA monotherapy became statistically significant (Figure 5 and Figure 6).

The effect of TOF plus cDMARD compared with cDMARD was bigger than GOL plus cDMARD, but smaller than non-TNFi, ETN, TNFi, RTX, TCZ and ABT in combination with cDMARD in the bDMARD-IR population using Estimate 2 (Figure 4). None of the treatment effects versus cDMARD were statistically significant, but the ERG suspects that a vague prior was used because the estimated between-study standard deviation was reported to have mean 1.21 with 95% credible interval (0.02, 4.52) which does not reflect the prior that the ERG has suggested. The company did not provide the results using Estimate 1.

The absolute treatment effects, including at least a moderate and at least a good EULAR response for both populations, are presented in Appendix 2.

A primary endpoint of radiographic progression using the mTSS in ORAL Scan was not significant at either 6 or 12 months (p=0.0792). Further statistically significant benefits for tofacitinib in combination with methotrexate (at 6 months) and for tofacitinib monotherapy (at 3 months) over placebo were observed using the EQ-5D, FACIT-F and pain assessed VAS outcomes ($p \leq 0.001$).

ACR20 at 3 months was significant for tofacitinib monotherapy versus placebo at 3 months in one trial (ORAL Solo) but not significant for the primary endpoint of the proportion achieving remission using DAS28(ESR) at 3 months. As all patients crossed over from placebo to receive tofacitinib at 3 months in ORAL Solo, there are no placebo-controlled results at 6 months for the other relevant endpoints. The ERG consider that the recently completed head-to-head trial, ORAL Strategy, has data relevant to the decision problem. The ORAL Strategy trial showed tofacitinib combination therapy with methotrexate to be non-inferior to adalimumab plus methotrexate but tofacitinib monotherapy was not found to be non-inferior to both tofacitinib plus methotrexate and adalimumab plus methotrexate for the primary endpoint of ACR50 at 6 months.

Safety data for tofacitinib were presented in the CS from a pooled analysis of tofacitinib trial data up to March 2015 which was two years prior to the current appraisal. Whilst the company were able to provide some up-to-date safety data following a request, the ERG note that a full and transparent safety profile of tofacitinib versus comparators, which contains comprehensive data for all AEs including SAEs, was not provided. The company stated that they were "*unable to update the incidence of Serious Adverse Events within the timelines provided as these are listed in a separate data base*". One of the most common AEs for tofacitinib was herpes zoster, which was also noted from a published NMA to be significantly higher than bDMARD comparators. [48] Incidence rates in the company's safety set were highest for serious infection events, bronchitis, pneumonia and all cardiac disorders. The ERG considers that pooling trials to produce incidence rates of AEs with tofacitinib may dilute the appearance of adverse events for tofacitinib plus cDMARD, which are noted by several sources[41, 49, 50, 114] to be higher than for tofacitinib monotherapy, which are not referenced or discussed in the CS. Moreover, the company's reliance on AE data from their own trial programme without performing targeted searches for relevant safety literature for tofacitinib means that relevant studies regarding safety, such as NMAs versus other bDMARDs, are missed.

The ERG believes that the results presented in NMA should be treated with caution, as the ordered categorical EULAR data were dichotomised in the cDMARD-IR population, which ignores the natural ordering and correlations between the EULAR response categories. A fixed effect model was used in all the analyses in the bDMARD-IR population and EULAR response (moderate response and good response) in the cDMARD-IR population. Heterogeneity is expected and this approach underestimates uncertainty in the treatment effect. For tofacitinib trials with early escape, the results

Table 13: Population characteristics at baseline used in the model

	cDMARD-IR		bDMARD-IR
	Moderate RA	**Severe RA**	**Severe RA**
Age	[redacted]	[redacted]	[redacted]
Proportion female	[redacted]	[redacted]	[redacted]
Weight (Kg)	[redacted]	[redacted]	[redacted]
HAQ-DI score	[redacted]	[redacted]	[redacted]
DAS28	[redacted]	[redacted]	[redacted]
Proportion with prior cDMARD experience	[redacted]	[redacted]	[redacted]
Proportion with prior bDMARD experience	[redacted]	[redacted]	[redacted]
Proportion anti-CCP positive	[redacted]	[redacted]	[redacted]
Disease duration (years)	[redacted]	[redacted]	[redacted]
Haemoglobin	[redacted]	[redacted]	[redacted]
CRP	[redacted]	[redacted]	[redacted]
ESR	[redacted]	[redacted]	[redacted]
Total cholesterol	[redacted]	[redacted]	[redacted]
CDAI	[redacted]	[redacted]	[redacted]
Number of previous DMARDs	[redacted]	[redacted]	[redacted]

Abbreviations: bDMARD, biologic disease-modifying anti-rheumatic drug; CCP, cyclic citrullinated peptide; CDAI, clinical disease activity index; cDMARD, conventional disease-modifying anti-rheumatic drug; CRP, c-reactive protein; DAS28, Disease Activity Score in 28 joints; ESR, erythrocyte sedimentation rate; HAQ, Health Assessment Questionnaire; IR, inadequate response.

5.3.2 Interventions and comparators

Descriptions of the intervention and the comparators are provided in Sections 3.2 and 3.3. Table 94 of the CS provides a summary matrix of which interventions are licenced (in combination with MTX or as monotherapy) in each of the moderate RA cDMARD-IR, moderate RA bDMARD-IR, severe RA cDMARD-IR, and severe RA bDMARD-IR populations. This table also includes information on recommendations provided by NICE. Table 34 summarises the comparators presented in the analyses within the CS. The ERG notes that some of the comparators included are currently not recommended by NICE and more importantly that recommended comparators are missing from some of the analyses presented by the company. The CS did not identify publications for inclusion of adalimumab, infliximab and certolizumab pegol for the bDMARD-IR populations. However, the ERG does not expect this to affect the conclusions of the company's economic analysis.

Table 14: ORs and probabilities of good and moderate EULAR response for each treatment used in the MTX-tolerant population

Therapy	ORs compared with TOF		Probabilities of EULAR response*		
	Moderate or good	Good	No response	Moderate response	Good response
TOF + MTX					
ADA + MTX					
CTZ + MTX					
ETN + MTX#					
ABT + MTX					
GOL + MTX					
IFX + MTX#					
RTX + MTX					
TCZ + MTX					
cDMARD†					

TOF: tofacitinib; ABT: abatacept; TCZ: tocilizumab; ADA: adalimumab; CTZ: certolizumab pegol; ETN: etanercept; GOL: golimumab; IFX: infliximab; RTX: rituximab; MTX: methotrexate; LEF: leflunomide; cDMARD: conventional disease-modifying antirheumatic drug

*Average probabilities based on the full population of ORAL trials (Scan, Standard, Sync and Step)

Biosimilars assumed to have same efficacy

†Includes MTX, LEF and cDMARD combination

Table 37 shows the ORs used in the model together with the average probabilities of moderate or good EULAR response for patients who could not tolerate MTX or for whom MTX was contraindicated. The probabilities of EULAR response for SSZ+HCQ were assumed to be equal to placebo. The ERG notes that this is likely to be an underestimate. Average probabilities were calculated averaging the probabilities of all patients in the ORAL Solo trial.

Table 15: ORs and probabilities of good and moderate EULAR response for each treatment used in the MTX-intolerant population

Therapy	ORs compared with TOF		Probabilities of EULAR response*		
	Moderate or good	Good	No response	Moderate response	Good response
TOF					
ADA					
ETN#					
TCZ					
SSZ+HCQ †					

TOF: tofacitinib; TCZ: tocilizumab; ADA: adalimumab; ETN: etanercept; GOL: golimumab; SSZ: sulfasalazine; HCQ: hydroxychloroquine

*Average probabilities based on the full population of ORAL Solo

Biosimilars assumed to have same efficacy

†Assumed equal to placebo

moderate, (iii) high and (iv) severe. Norton *et al*. report a regression model to calculate each patient's probability of belonging to each class based on the patient's baseline characteristics. The company follow the approach used by the AG in TA375 whereby the change in HAQ-DI score for a patient is calculated as the weighted change in HAQ-DI associated with each class. The company provides commercial-in-confidence data that show that the patients in the ORAL trials are more likely to be in a worse HAQ-DI progression class than the ERAS cohort[121] and that assumed within TA375.[24] This may be due to the recruitment of patients with established RA in the ORAL trials.

In the second approach, the company assumed that 'rapid progressors' could be identified. These patients are assumed to have a worse long-term HAQ-DI prognosis than that for average patients, which was taken from work reported by the NICE Decision Support Unit (DSU).[122] The ERG comments that whether such patients could be identified has been questioned in a report by Stevenson *et al.*[123] considered within TA375. Furthermore, the company producing baricitinib, having analysed academic-in-confidence data on changes in HAQ, stated in its submission to NICE that '*this suggests that the 'rapid-progressor' group discussed in TA375 that might benefit from more aggressive treatment is a small minority of the overall moderate population*.'[124]

An additional scenario analysis was performed that assumed that HAQ-DI progression was linear for patients receiving cDMARDs and that HAQ-DI increased at a rate of 0.045 per year for patients on LEF and at a rate of 0.06 per year for patients on PALL. The ERG believes that these analyses are inappropriate as HAQ-DI progression has been proven to be non-linear[122] in TA375.[24]

HAQ-DI trajectory prior to treatment cessation

The CS states that prior to treatment discontinuation, the HAQ-DI score improvement observed upon treatment response was lost linearly over the six-month period. This is similar to the approach used in TA375,[24] although in TA375 the entire HAQ-DI loss occurred at the time of discontinuation.

After applying changes to HAQ-DI scores, the resulting values were rounded to the nearest valid HAQ-DI score (which is a multiple of 0.125). The ERG notes that this approach can lead to inaccurate results. This contrasts with the approach used in TA375[24] in which scores were rounded to either the higher or the lower valid HAQ-DI score with a probability proportional to their distance to each (e.g. a value twice closer to the upper HAQ-DI score would be twice as likely to be simulated as the upper score than simulated as the lower score). This point was raised by the ERG during the clarification process (see clarification response,[34] question B4) but was misunderstood and therefore not addressed by the company despite the code being contained in the model to perform a probabilistic analysis of HAQ-DI changes. The ERG assessed the impact of this change in its exploratory analyses.

1. Limitations with the company's NMA
2. Missing comparators
3. Inadequate sequences of treatments
4. Assuming same efficacy for SSZ+HCQ as for placebo
5. Assuming the efficacy of the first bDMARD applies to all treatment lines of bDMARDs in the cDMARD-IR population
6. Assuming the same efficacy for TOF+MTX and TOF monotherapy
7. Deterministic rounding to nearest HAQ-DI score
8. Linear HAQ-DI trajectory for palliative care

1. Limitations with the company's NMA

The ERG believes that the company's NMA suffers from potential limitations, which have been described in Section 4.4: (i) the ordered categorical EULAR data were dichotomised in the cDMARD-IR population, which ignores the natural ordering and correlations between the EULAR response categories; (ii) a fixed effects model was used in all the analyses in the bDMARD-IR population and for EULAR responses, which underestimates uncertainty in the treatment effect; and, (iii) the imputation approach used in TOF trials potentially overestimates the treatment effect of TOF versus cDMARD, and could have an important impact in the position of TOF among the bDMARDs.

2. Missing comparators

The company's analyses did not include all the relevant comparators for some of the populations as explained in Section 5.2.3 and Table 34. Most importantly, all relevant comparators were missing in the analysis for bDMARD-IR MTX-intolerant patients with severe RA and four comparators (ADA, ETN, IFX and CTZ with concomitant MTX) out of seven were missing from the analysis for bDMARD-IR RTX-ineligible patients with severe RA. The CS did not identify publications for inclusion of adalimumab, infliximab and certolizumab pegol for the bDMARD-IR populations. The ERG notes the company included neither the RTX biosimilar nor the SC formulations of ABT and TCZ.

3. Inadequate sequences of treatments

The ERG notes that the sequences used by the company were not appropriate for the following reasons:

- The inclusion of multiple consecutive treatments of cDMARD combinations and SSZ+HCQ. Patients only go through one such treatment before progressing to another type of treatment.
- The inclusion of bDMARD treatments in populations and points in the pathway which have not been recommended by NICE, such as:
    - ETN+MTX after TCZ+MTX and RTX+MTX in cDMARD-IR patients with severe RA.

- ABT+MTX and GOL+MTX in the bDMARD-IR RTX-eligible patients with severe RA.
- TCZ+MTX after TOF, ABT or GOL concomitant with MTX in the bDMARD-IR RTX-ineligible patients with severe RA.
- GOL+MTX after TCZ+MTX in the bDMARD-IR RTX-ineligible patients with severe RA.
- TCZ monotherapy in bDMARD-IR MTX-intolerant patients with severe RA.
- RTX+MTX and TCZ+MTX after cDMARD combination in cDMARD-IR patients with moderate RA.

- The inclusion of three or four post-biologic treatments before palliative care instead of just one.

4. *Assuming the same efficacy for SSZ as for placebo*

The company used the EULAR response ORs calculated in the NMA for placebo as an estimate for the ORs for SSZ+HCQ. The ERG notes that this is likely to underestimate the effectiveness of SSZ and therefore underestimate the ICER for TOF monotherapy compared with SSZ.

5. *Assuming the same efficacy for TOF as monotherapy and in combination with MTX*

The company assumed that TOF as monotherapy would have the same efficacy as in combination with MTX. However, ORAL Strategy (NCT02187055)[40] showed that TOF monotherapy was not found to be non-inferior to TOF+MTX. The also NMA shows that TOF monotherapy results in slightly lower probabilities of response than TOF + MTX: in cDMARD-IR patients, an average of [redacted] versus [redacted] achieved good EULAR response and [redacted] versus [redacted] achieved moderate EULAR response (see clarification response,[34] Table 8). However, the ERG acknowledges that the company estimated the efficacies of other monotherapies in comparison with TOF monotherapy and therefore the relative impact of this assumption is likely to be reduced.

6. *Assuming the efficacy of the first bDMARD applies to all treatment lines of bDMARDs in the cDMARD-IR population*

Within the CS, the company assumed that the efficacy of bDMARDs in terms of probabilities of EULAR response would remain unchanged irrespective of whether they were given as first line or subsequent line treatment. However, as demonstrated by the company's own regression model, the efficacy of bDMARDs is lower in bDMARD-IR patients than in cDMARD-IR patients. Therefore, for the second and subsequent lines of treatment in the cDMARD-IR population, it is more appropriate to use the probability of EULAR response calculated in the bDMARD-IR patients. During the clarification process, the ERG asked the company to activate

the prior_bdmard flag after patients had gone through their first bDMARD (or JAK inhibitor). The company implemented this change and