CONFIDENTIAL UNTIL PUBLISHED
Evidence Review Group's Report
Tocilizumab for treating giant cell arteritis

Produced by	CRD and CHE Technology Assessment Group, University of York, Heslington, York YO10 5DD
Authors	Sahar Sharif, Research Fellow, CRD
	Simon Walker, Research Fellow, CHE
	Nerys Woolacott, Reader in Health Technology Assessment, CRD
	Kath Wright, Information Specialist, CRD
	Stephen Palmer, Professor of Health Economics, CHE
Correspondence to	Stephen Palmer, Professor of Health Economics, CHE
Date completed	Date completed (12/10/2017)

Source of funding

This report was commissioned by the NIHR HTA Programme as project number 13/134/03.

Declared competing interests of the authors

None

Acknowledgements

We would like to thank Dr Sarah Mackie, Consultant Rheumatologist, Leeds Teaching Hospitals, NHS Trust and Associate Clinical Professor in Vascular Rheumatology, University of Leeds, for her advice on the management of Giant Cell Arteritis. Dr Mackie has in the past acted as an advisor to Roche in relation to and was also a sub-investigator (listed on delegation log) on the Roche sponsored clinical trial, GiACTA. She has also been a site investigator for a trial of a competitor treatment in Giant Cell Arteritis.

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:

Sharif S, Walker S, Woolacott N, Wright K, Palmer S. Tocilizumab for treating giant cell arteritis: A Single Technology Appraisal. CRD/CHE University of York, 2017.

Contributions of authors

Please refer to the International Committee of Medical Journal Editors (ICMJE) Uniform Requirements for Manuscripts Submitted to Biomedical Journals see http://www.icmje.org/

Note on the text

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

Copyright statement

Table of Contents

Table of Tables

Table of Figures

List of abbreviations

ACR	American College of Rheumatology
AE	Adverse event
AFT	Accelerated failure time
ALT	Alanine transaminase
ACT	Aspartate transaminase
AESI	Adverse events of special interest
BHPR	British Health Professionals in Rheumatology
BSR	British Society for Rheumatology
CE	Cost-effectiveness
CEAC	Cost-effectiveness acceptability curve
CHMP	Committee for Medicinal Products for Human Use
CI	Confidence interval
CPRD	Clinical Practice Research Datalink
CRP	C-reactive protein
CS	Company submission
CSR	Clinical study report
CUA	Cost utility analysis
DH	Department of Health
DSU	Decision Support Unit
EE	Economic evaluations
EMA	European Medicines Agency
EPAR	European Public Assessment Report
EQ-5D	EuroQoL 5 Dimensions questionnaire
EQ-VAS	EuroQoL visual analogue scale
ERG	Evidence Review Group
ESR	Erythrocyte sedimentation rate
FACIT	Functional Assessment of Chronic Illness Therapy
FDA	(US) Food and Drug Administration
FTP	Fast Track Pathway
FU	Follow-up
GC	Glucocorticoid
GCA	Giant Cell Arteritis
HCRU	Healthcare Resource Utilisation
HE&OR	Health Economics and Outcomes Research
HIRU	Health Information Research Unit

HR	Hazard ratio
HRQL	Health-related quality of fife
HSUV	Health-state utility values
HTA	Health Technology Assessment
HUI	Health utility index
ICD-9-CM	International Classification of Diseases, Clinical Modification
ICER	Incremental cost-effectiveness ratio
Incr	Incremental
IQR	Interquartile range
ITT	Intention-to-treat
IU	International Units
IV	Intravenous
IVRS	Interactive Voice-Response System
KM	Kaplan-Meier
LTFU	Long-term follow-up
LSM	Least square means
LV	Large vessel
LYG	Life years gained
MCDA	Multi criteria decision analysis
MCS	Mental Component Score
MEDLINE	Medical Literature Analysis and Retrieval System Online
MHRA	Medicines and Healthcare products Regulatory Agency
MR	Magnetic resonance
MRU	Medical Resource Utilisation
MTX	Methotrexate
NA	Not applicable
NE	Not evaluable
NHS	National Health Service
NHS EED	NHS Evidence Evaluation Database
NICE	National Institute for Health and Care Excellence
NR	Not reported
NS	Not significant
OCS	Oral corticosteroids
OLE	Open-label extension
ONS	Office for National Statistics
P	p-value
PAS	Patient access scheme

PASLU	Patient access scheme liaison unit
PBO	Placebo
PBO +26	Placebo plus a 26 week glucocorticoid taper
PBO + 52	Placebo plus a 52 week glucocorticoid taper
PCS	Physical Component Score
PFC	Points for clarification
PGA	Patient global assessment
PH	Proportional hazards
PIM	Promising Innovative Medicine
PK	Pharmacokinetics
PMR	Polymyalgia rheumatica
PRISMA	Preferred reporting items for systematic reviews and meta-analyses
PRO	Patient-reported outcomes
PSA	Probabilistic sensitivity analysis
PSS	Personal Social Services
PY	Patient years
QALY	Quality-adjusted life years
QW	Once per week
Q2W	Once every other week
RCT	Randomised, controlled trial
RA	Rheumatoid arthiritis
SAE	Serious adverse events
SAS	Statistical Analysis System
SC	Subcutaneous
SD	Standard deviation
SE	Standard error
SF	Short Form
SIGN	Scottish Intercollegiate Guidelines Network
SLR	Systematic literature review
SMC	Scottish Medicines Consortium
SMQN	Standardised MedDRA Query, narrow (scope)
SOC	System organ class
STA	Single technology appraisal
TAB	Temporal artery biopsy
TB	Tuberculosis
TCZ	Tocilizumab
TCZ QW +26	Tocilizumab once a week plus a 26 week glucocorticoid taper

TCZ Q2W +26	Tocilizumab twice a week plus a 26 week glucocorticoid taper
TEAE	Treatment-emergent adverse event
TNF	Tumour necrosis factor
TTFF	Time to first flare
TTO	Time trade off
UK	United Kingdom
ULN	Upper limit of normal
US	Ultra sound
VAS	Visual analogue scale
VBA	Visual Basic for Applications
WHO	World Health Organisation

1 Summary

Giant Cell Arteritis (GCA) is an inflammatory vasculopathy affecting large and medium-sized arteries. The company submission (CS) stated that GCA is a potentially life-threatening condition linked with substantial impairment of the day-to-day functioning of patients. The ERG believes that describing GCA as a potentially life threatening condition is not well substantiated: whilst GCA may rarely lead to life threatening events such as aortic aneurysm rupture or stroke, at a population level there is no clear evidence that long-term mortality is significantly increased in patients with GCA compared to individuals without GCA. The CS describes two clinical subtypes of GCA: cranial GCA which is the most typical presentation; and large vessel (LV) GCA which is less common. Cranial GCA can result in ischaemic manifestations such as severe headache, jaw claudication and visual impairment. Clinical advice to the ERG indicated that once treatment is initiated it is rare for patients to develop vision loss. The CS describes the complications of LV GCA as aortic aneurysms, aortic dissection and coronary arteritis.

GCA is a rare condition, it is estimated that around 1 in every 4,500 people will develop it in the UK each year. The CS stated that GCA primarily affects adults ≥50 years old. The risk increases with age, with the highest rates being observed between 70 and 80 years. The CS correctly stated that there are no NICE guidelines for GCA; however, the British Society for Rheumatology (BSR) has developed clinical practice guidelines to advise the diagnosis and management of GCA. The intervention presented is tocilizumab (TCZ), which is currently awaiting marketing authorisation, expected in September 2017.

The CS reports that current treatment mainly consists of high dose GC (usually prednisone – the ERG notes that in the UK this is usually prednisolone) followed by long-term steroid tapering. Complicated GCA (evolving vision loss or established vision loss) is treated with an initial dose of 60 mg or above, whereas uncomplicated GCA (no jaw or tongue claudication or visual symptoms) is treated with 40-60 mg. Once signs and symptoms of GCA are absent patients are slowly tapered off GC.

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

The population for this submission were adults with GCA, which was in line with the NICE scope definition. The ERG clinical advisor stated that the GiACTA trial population was generally applicable to patients seen in NHS practice, with the exception that there were a higher proportion of patients with large vessel GCA, than is typically seen in NHS practice.

The intervention presented in the CS was tocilizumab, which matched that specified in the NICE scope. The recommended posology is 162 mg of subcutaneous tocilizumab once every week in combination with a tapering course of GC. In the GiACTA trial there were two tocilizumab arms:

once a week (QW) dosing and one every other week (Q2W) dosing; only the once a week dosing is licensed and therefore, this report presents tocilizumab results for this dose only. The GC taper used alongside tocilizumab lasts 26 weeks. The ERG notes that this is much shorter than the length of GC taper used in current clinical practice (see further discussion of this below). Although it is likely that a 26 week taper would be attempted with tocilizumab in practice, with the aim of gaining the potential steroid sparing benefits of tocilizumab, it is not certain how generally this would be achieved.

The comparator for this submission was established clinical management without tocilizumab. The comparator used in the GiACTA trial was placebo with either a short (26 weeks) or long (52 weeks) prednisone taper regimen according to a defined schedule. This matched the NICE scope. The CS clarified that prednisone/prednisolone was used as it is the mainstay of treatment for people with GCA; published evidence and clinical advice to the ERG confirmed that in the NHS prednisolone is used rather than prednisone. The ERG notes that prednisolone and prednisone are highly comparable drugs, prednisone being the metabolic precursor of prednisolone. The GiACTA trial used two different placebo controls: one with a 26 week GC taper and one with a 52 week taper. The ERG notes that the BSR recommends a GC tapering regimen which adds up to a minimum of 52 weeks and a cumulative GC dose between 3.6g and 7.4g over approximately 1 - 1.5 years, in those patients who do not experience a relapse or flare. Therefore, the placebo+52 week GC taper is the more relevant comparator for UK clinical practice.

The outcomes measures for the submission were: disease remission, time to relapse after disease remission, GC exposure, adverse effects of treatment and health-related quality of life. These essentially matched the outcomes listed in the NICE scope. However, morbidity (including vision loss) was listed in the NICE scope as an outcome but vision loss was not reported in the CS as a separate outcome. After this issue was raised in the points for clarification, the company confirmed that vision loss was recorded as part of the clinical assessment for each patient at each study visit. The company pointed out that, “The level of clinical excellence employed by the investigators in monitoring disease activity ensured that any increase in disease activity was appropriately treated to prevent severe complications such as permanent vision loss.” Therefore, the ERG agrees that vision loss is not an important trial outcome.

1.2 Summary of clinical effectiveness evidence submitted by the company

The CS presented two RCTs of tocilizumab in GCA: the Phase II study (NCT01450137) and a Phase III RCT (GiACTA). GiACTA is the only RCT that provides data on the effectiveness of the licensed dose of tocilizumab in patients with GCA. The Phase II study (NCT01450137) provides only supporting evidence. The CS stated it would not be appropriate to attempt pooling of the efficacy data from the Phase III GiACTA study and the Phase II NCT01450137 study because of differences in treatment regimens and study designs, therefore a standard meta-analysis was not feasible.

The GiACTA trial investigated the clinical effectiveness of tocilizumab in 251 adults over 50 years old (mean age 69 years) with new-onset or relapsing giant cell arteritis. The trial consisted of four arms, however this report focuses on the arms most applicable to UK clinical practice: 162mg of tocilizumab once a week with a 26 week GC taper (TCZ QW+26) (n=100) and placebo with a 52-week GC taper (PBO+52) (n=51). The tocilizumab treatment duration was 52 weeks.

Sustained remission

Tocilizumab was more effective than placebo in sustaining remission, with a significantly higher number of participants with sustained remission at Week 52 in the TCZ QW+26 arm (56.0%) compared with the PBO+52 arm (17.6%); the difference in percentage of responders was 38.35 (99% CI 17.89 to 58.81) ($p<0.0001$).

The GiACTA trial has an ongoing Part 2, which is an open-label extension including patients from Part 1 who will be followed for an additional 2 years. Preliminary results from Part 2 were that 33% of TCZ QW+26 responders flared after discontinuation of tocilizumab, indicating that for a sustained treatment benefit, continued treatment with tocilizumab is needed in a substantial proportion of patients. Therefore, further reliable and accurate research is needed to determine the long term effectiveness of tocilizumab in maintaining remission in patients with GCA.

Flare

The hazard ratio (0.37, 99% CI: 0.2-0.7) showed a statistically significant lower risk of flare in patients in the tocilizumab group compared to the placebo+52 week group ($p<0.0001$). The mean annualised relapse rate for multiple flares observed in each patient was 1.30/year in the PBO+52 arm (median: 1) compared with 0.41/year in the TCZ QW+26 arm (median:0).

Cumulative dose of GC

There was a statistically significant lower median cumulative GC dose to Week 52 in the TCZ QW+26 group (1862mg) when compared with the PBO+52 group (3817.5mg) ($p<0.0001$).

Sub-group analyses

Sub-group analyses by disease status at baseline (new-onset or relapsing) for Sustained Remission at week 52, for Time to GCA flare, and for cumulative GC dose were reported in the CS.

The difference in the proportion of patients achieving sustained remission at Week 52 between the TCZ QW+26 group and the PBO+52 group was similar among new-onset (37.9%) and relapsing GCA patients (38.5%). However, the proportion of patients in sustained remission in the PBO+52 group was lower for relapsing patients (14.3%) than for new-onset patients (21.7%).

The median time to GCA disease flare in new-onset GCA patients was 169 days in the PBO+26 group and was not calculable for the other three groups due to fewer than 50% of the new-onset patients in these groups experiencing a flare. In relapsing patients it was 165 days in the PBO+26 group and 274 days in the PBO+52 group but was not calculable in the tocilizumab treatment groups. The ERG analysed both subgroups and found that the relative treatment effect was slightly less in the new-onset patients (HR 0.44, 95% CI 0.29 -1.59; (p=0.004)) compared with the relapsing patients (HR 0.33, 95% CI 0.14 – 0.81; (p=0.04)

The mean differences between cumulative dose in the TCZ QW arm and the PBO+52 arm for these subgroups were not compared formally, but it was numerically higher in the relapsing patients (2426 mg compared with 1730 mg) despite their lower GC dose at baseline.

Health related quality of life

There was no notable deterioration observed in HRQL in any treatment group, however the tocilizumab groups appeared to score marginally better. The only statistically significant difference was seen for the SF-36 Physical Component Score. There was no substantial deterioration in the EQ-5D scores in any treatment group. Numerically higher mean changes in the FACIT-F from baseline were observed for the tocilizumab treatment group versus the placebo group. However, no statistical testing was performed. Both, the TCZ QW+26 and PBO+52 groups showed a numerical improvement from baseline in the Mental Component Score; however there was no significant difference. Therefore, there is limited evidence to indicate that HRQL improves substantially with tocilizumab compared to placebo.

Adverse effects of tocilizumab

The safety profile of tocilizumab appears to be similar to the placebo used in the trial. The total number of patients with at least one AE was similar across all treatment groups; however it was highest in the TCZ-QW group (98.0%) and lowest in the PBO+52-week group (92.2%). Furthermore, there were a higher number of patients experiencing infections in the TCZ QW+26 group (75%) compared with the PBO+52 group (64.7%) (Table 22, Page 64 of the CS). As tocilizumab is given with the intention of being steroid sparing it might be hoped that GC-associated AEs would be lower in the TCZ QW+26 arm. In GiACTA however, the percentage of patients reporting an AE considered related to GC use by the investigator was similar in the TCZ QW+26 (50%) and PBO+52 (49%) groups.

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

The GiACTA trial was a large, relatively good quality, double-blinded, RCT. However, there were some prognostic factors which were unbalanced between the four arms in the GiACTA trial: these imbalances may slightly reduce the reliability of the study results. In addition, the primary outcome of

sustained remission at Week 52 and the secondary outcome of time to first GCA flare may be biased due to not all patients being in remission at baseline. The chance of a placebo patient, who was not in remission at baseline, achieving remission at week 12 may be biased against by the imposition of the GC taper from baseline. In contrast, the time of first GCA flare may be biased in favour of placebo due to not all patients being in remission at baseline.

The generalisability of the GiACTA trial to the UK GCA population is generally appropriate, however there are some differences:

- The number of patients from the UK in the TCZ QW+26 arm of the trial was only 7.
- The GiACTA trial includes both new-onset and relapsing GCA patients. Clinical advice to the ERG indicated that these two subgroups of patients would be treated differently in practice. The analysis of the GIACTA trial can be criticised because it did not take into account the difference between new-onset and relapsing patients, nor that between those who were in remission at baseline and those who were not. Randomisation was stratified by baseline prednisone dose only. Whilst there was a significant difference in baseline prednisone dose between new-onset and relapsing patients, this stratification will not account for the other differences between the new-onset and relapsing populations. Sub-group analyses by disease status at baseline (new-onset or relapsing) for sustained remission at week 52, for time to GCA flare, and cumulative GC dose were reported in the CS.
- The baseline characteristics of the GiACTA population appear to be fairly representative of the UK GCA population. However, the ERG notes that there is a difference in the mean age of patients in the GiACTA trial (69.05 years) and that from the UK CPRD data source (73 years). Also, overall there was a higher ratio of large vessel GCA patients to cranial GCA patients than would be seen in NHS practice.
- The trial uses a 26 week GC taper for three of the four treatment groups. The tapering regimen recommended by BSR adds up to a minimum of 52 weeks. Hence, the placebo arm with a 52 week GC taper is most relevant to UK clinical practice. The 26 week taper used with tocilizumab is likely to be attempted in clinical practice, with the aim of reducing the GC load.
- Although the trial included four treatment arms the only comparison relevant to NHS practice is that between TCZ+26 and PBO+52

1.4 Summary of cost effectiveness submitted evidence by the company

The company's economic submission included a systematic review of published evidence on the cost-effectiveness of tocilizumab for GCA and a separate model. The review identified a single previously

published study that assessed the lifetime costs and consequences of two tocilizumab doses (TCZ QW and TCZ Q2W) in combination with a 26 week prednisone taper regimen compared to a 52 week prednisone taper regimen alone. The published study shares an identical structure and many common inputs and assumptions to the company model. The ERG considered that the cost-effectiveness analysis reported in the company model to be the most relevant source of evidence to inform the decision problem.

The company submission was based on a semi-Markov model using a weekly cycle length. The model evaluated the lifetime (30-years) cost-effectiveness of tocilizumab in combination with a 26-week prednisone taper regimen compared to a 52-week prednisone taper regimen alone. The model used GiACTA trial data to estimate the impact of tocilizumab on disease control (e.g. time in remission and number of flares) and real world data to estimate the effect of steroid sparing. The real world data was used to quantify the relationship between cumulative prednisone dose and the risk of steroid related adverse events in GCA patients.

The model included seven separate health states: (i) On remission and on steroid; (ii) On remission and off steroid; (iii) On relapse/flare; (iv) On remission and on maintenance steroids (escape); (v) GCA-related complications; (vi) Steroid-related AEs and (vii) Death.

Separate remission states were used before and after a first flare to account for different transition probabilities and glucocorticoid (GC) exposure based on GiACTA trial data. GCA-related complications (vision loss and stroke) were assumed to only occur from the relapse/flare state and transitions were derived from external literature. Steroid-related AEs included fractures and diabetes based on cumulative GC dose and evidence from real world data. Death included background mortality (general population, age and gender matches) arising from any state with an adjustment for stroke related mortality attributed to GCA-related complications

Treatment with tocilizumab was assumed to be continued over a 2-year fixed treatment period in the base-case analysis .This was justified by the company based on the CHMP Positive Opinion which states that tocilizumab can be continued beyond 1-year, clinical opinion and the typical duration of conventional treatment for GCA with GCs. The 52-week GC tapering regimen included in the GiACTA trial was considered an appropriate comparator and consistent with the most rapid GC tapering regimen recommended in clinical guidelines.

Transition probabilities from the initial remission state to the first flare/relapse event were based on individually fitted parametric models using patient-level data from the ITT population of the GiACTA trial. Transition probabilities from the subsequent remission state to flare were based on a separate

Poisson regression. The effectiveness of tocilizumab was assumed to be maintained over a lifetime and justified by the company based on early results from open label data.

The risk of GCA related complications was assumed to be related to subsequent relapses/flares. In the absence of these complications arising in the GiACTA trial, estimates were sourced from a separate published economic model comparing alternative diagnostic approaches for GCA. The use of external evidence was justified by the company as these events are rare but associated with significant and potentially lifelong cost and health-related quality of life (HRQoL) implications.

Cumulative GC dose for each treatment arm were based on 3 separate estimates to reflect dosing during: (i) the initial remission period (prior to first flare), (ii) during secondary remission (post-initial flare) and (iii) during relapse/flare. Dose estimates were based on data from the GiACTA trial and real world data. Background mortality was derived from standard lifetables and justified based on findings from a systematic review which found no significant differences in mortality for GCA patients.

Utilities for the remission and relapse/flare states were sourced from a mixed effect regression model based on EQ-5D data from GiACTA. Data was combined across the separate arms given the lack of significant difference by treatment arm reported within the trial. The relapse/flare utility was applied for a 4-week duration based on published literature and clinical opinion. Utility decrements for GCA and GC-related complications were sourced from the external literature.

The treatment costs of tocilizumab and GC were based on published prices and the approved PAS scheme for tocilizumab. No additional administration costs were assumed for tocilizumab. The cost of conventional GC treatment was based on published prices for prednisone. Following points for clarification, the company altered the costs for conventional GC treatment using published prices for prednisolone which is more commonly used in the NHS. Health state costs were based on third-party market research undertaken by the company. The costs of GCA related complications and GC related AEs were derived from the external literature.

The company's base-case results were based on the overall ITT population. Separate results for the subgroups identified within the NICE scope (newly diagnosed and relapsed/refractory) were included in the company's response to the points for clarification. Only the results from the PAS analysis were considered by the ERG.

The company base-case deterministic ICER for tociluzumab treatment with GC versus GC alone for the ITT population was £28,272 per additional QALY. The subgroups ICER's were £37,334 per QALY in the newly diagnosed subgroup and £22,403 per QALY in the relapsed/refractory subgroup.

The disaggregated QALY data provided by the company showed that the main driver of incremental QALY gains was the additional time patients are assumed to be in one of the remission states with tocilizumab treatment. The QALY gains are derived from two main sources: (i) a longer time to first flare, which means that patients receive the higher utility of remission and avoid the utility decrement of GC-related AEs; (ii) fewer subsequent relapse/flare events. The impact of differences due to GCA-related complications was minor.

The disaggregated cost data indicated that the main driver of cost differences was the additional acquisition cost of tocilizumab treatment. These additional costs were partially offset by a lower disease management cost (i.e. longer time in the 'On remission and off steroid state) and reduced flare costs. Additional cost-offsets were assumed in terms of reduced GCA-complications and GC-related adverse events. However, these offsets appeared less significant than the disease management and flare costs.

The major driver of the differences in the ICER estimates across the populations was differences in the total number of flares. The incremental difference in the number of flares was estimated to be -5.87 in the newly diagnosed and -19.21 in the relapsed/refractory subgroups, compared to -12.24 in the base-case ITT population. The differences across the different populations were due to different parametric functions for the time to first flare and different rates of subsequent relapse/flare events.

The probabilistic base-case ICER reported by the company for the ITT population was £30,579 per QALY. The ERG was unable to replicate the company probabilistic ICER estimates.

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

The ERG's logical checks identified an important error in the QALY calculations which was corrected by the company and a revised model and full set of results were provided by the company. Although the ERG was satisfied with the internal validity of the revised model, significant concerns remained regarding the clinical and external validity of the longer-term extrapolations and the extent to which the company model appropriately represented the natural history of GCA.

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

1.6.1 Strengths

Clinical

The systematic review conducted to identify relevant trials used methods that were generally appropriate; it is unlikely that any relevant randomised controlled trial (RCT) of tocilizumab has been missed. The CS presented two RCTs of tocilizumab in GCA: the Phase II study (NCT01450137) and a Phase III RCT (GiACTA). GiACTA is a good quality RCT that provides data on the effectiveness

of the licensed dose of tocilizumab in patients with GCA. The Phase II study (NCT01450137) provides supporting evidence.

Cost-effectiveness

The company's economic submission met the requirements of the NICE reference case. The company submission acknowledged many of the key uncertainties and the cost-effectiveness model incorporated a range of scenario analyses that allowed the impact of alternative assumptions to be explored. The company provided a revised model and included subgroups within their response.

1.6.2 Weaknesses and areas of uncertainty

Clinical

The treatment effect in new-onset vs relapsing patients was not fully explored, nor was the effect in patients with GCA vs LV or both.

The generalisability of the trial is uncertain due to the age of patients, the ratio of cranial vs LV GCA patients, and the uncertainty regarding the taper that will be used with tocilizumab in practice

The available preliminary evidence indicates that around 30% of patients will flare once tocilizumab treatment is stopped: for sustained treatment benefit, continued treatment with tocilizumab is needed in a substantial proportion of patients. Therefore, further reliable and accurate research is needed to determine the long term effectiveness of tocilizumab in maintaining remission in patients with GCA.

Cost-effectiveness

The ERG was concerned that the assumption that the benefits of tocilizumab continue over a lifetime regardless of the treatment duration did not appear to be justifiable based on early results from the OLE study and the published results from the previous RCT. The external evidence identified by the ERG also raised uncertainties regarding the external validity of the longer-term predictions.

1.7 Summary of exploratory and sensitivity analyses undertaken by the ERG

A series of additional revisions and alternative assumptions were explored by the ERG using additional scenarios. These scenarios explored uncertainties related to: (i) the duration of treatment and the assumption that the benefits of tocilizumab continue over a lifetime; (ii) uncertainty concerning the choice of parametric survival models for time to first flare and use of different model types and (iii) uncertainty concerning the rate of subsequent relapse/flares following an initial flare. The ERG proposed alternative assumptions and data sources which they considered had greater face validity and were more consistent with the natural history of GCA reported in longer-term epidemiological studies. These alternative approaches and data sources were then combined as part of an alternative ERG base-case analysis.

The ERG's alternative base-case presented results for alternative treatment duration periods between 1 and 2 years. The ERG ICER results were higher than those reported by the company. The ERG probabilistic ICERs for a 1-year treatment period were: £36,960 (ITT population); £41,577 (newly diagnosed subgroup) and £30,158 (relapsed-refractory subgroup) per QALY. The ERG probabilistic ICERs for a 2-year treatment period were: £65,801(ITT population); £73,046 (newly diagnosed subgroup) and £58,411 (relapsed-refractory subgroup) per QALY.

The ERG considers that the 1-year treatment period results provide the most internally valid estimates consistent with the treatment duration period assessed in the GiACTA trial. However, in the absence of a clear stopping rule for tocilizumab there remains significant uncertainty concerning the appropriate duration of tocilizumab treatment. The differences reported between the company and ERG highlight that important uncertainties remain concerning the optimal duration of tocilizumab treatment and the associated longer-term benefits.

2 Background

2.1 Critique of company's description of underlying health problem

The relevant health problem in the present appraisal is Giant Cell Arteritis (GCA), which is an inflammatory vasculopathy affecting large and medium-sized arteries, primarily the extracranial branches of the carotid arteries and the aorta's primary branches. The company submission (CS) stated that GCA is a potentially life-threatening condition linked with substantial impairment of the day-to-day functioning of patients.[1] The ERG believes that describing GCA as a potentially life threatening condition is not well substantiated: whilst GCA may rarely lead to life threatening events such as aortic aneurysm rupture or stroke, at a population level there is no clear evidence that long-term mortality is significantly increased in patients with GCA compared to individuals without GCA.[2] The CS reports mortality as an outcome; however it is not the main concern for GCA patients.[1] Although, the overall life expectancy of patients with GCA is similar to that of the general population, GCA can increase the risk of developing serious problems, debilitating patients and reducing their quality of life.[3, 4] The greatest driver of treatment decisions, for many doctors and patients, is most likely the fear of visual loss balanced against awareness of the burden of glucocorticoid therapy.[5]

GCA is a rare condition, it is estimated that around 1 in every 4,500 people will develop it in the UK each year.[3] The CS stated that GCA primarily affects adults ≥50 years. The risk increases with age, with the highest rates being observed between 70 and 79 years.[6, 7] The ERG requested the UK incidences of GCA in people aged over 50 and 70 years old as the CS did not initially provide these. The CS response stated that Petri et al.[7] reported the incidence in women aged 70-79 years old as 7.4 per 10,000 person-years, with an estimate of 3.7 per 10,000 years in men. The CS stated that the incidence in men peaks at age 80, whereas in women it peaks at age 70 to 79 years. The ERG notes that GCA is three times more common in women than in men and seven times more common in white people than in black people[8]; this was not stated in the CS.

The CS describes two clinical subtypes of GCA: cranial GCA which is the most typical presentation; and large vessel GCA which is less common. Clinical advice to the ERG is that these are two manifestations of the same disease, and that with increasing use of vascular imaging these two clinical subtypes may often be seen together in the same patient. Cranial GCA involves the extracranial branches of the carotid arteries and can result in ischaemic manifestations such as severe headache, scalp tenderness and jaw claudication. Serious manifestations/complications of cranial GCA relate to vision; these range from transient diplopia to sudden, partial or complete vision loss. The serious complication of vision loss usually manifests before or shortly after diagnosis and once established it is almost always permanent, but it can be prevented with early treatment.[9] Clinical advice to the ERG indicates that once treatment is initiated and appropriately managed, it is rare for patients to develop vision loss. Approximately 20% of untreated GCA patients have manifestations of vision loss,

whereas permanent vision loss can affect approximately 12-15% of patients.[6, 9, 10] The CS states that approximately 30% of GCA patients experience visual manifestations, but the ERG notes that this figure does not apply to actual vision loss.

Large-vessel GCA (LV GCA) affects the aorta and its primary branches. The CS describes the complications of LV GCA as aortic aneurysms, aortic dissection[11] and coronary arteritis.[12] Compared to the general population, aortic aneurysms are 17 times more likely in GCA sufferers. Most patients with GCA will develop aortitis but it manifests clinically in approximately 15% of patients.[12] Clinical advice to the ERG indicated that patients with large vessel GCA tend to have longer disease duration with more relapses, whereas patients with cranial GCA generally have shorter disease duration of approximately one to two years with fewer relapses compared to those with LV-GCA. A study by Alba et al. reported that a relapsing GCA course is associated with higher and prolonged GC requirements[13]. For this reason, patients with large vessel GCA are typically harder to treat compared to patients with cranial GCA.

The CS states that in approximately 90% of patients the erythrocyte sedimentation rate (ESR) or C-reactive protein (CRP) is elevated. However, clinical advice to the ERG puts this figure at 95% to 98%. Furthermore, CRP has been shown to be more effective in diagnosing GCA than ESR.[14]

Overall, the ERG believes that the CS generally presented appropriate and relevant information on the underlying health problem. However, the CS overstated the incidence of visual manifestations in patients with GCA and describing GCA as a life threatening condition was unsubstantiated.

2.2 Critique of company's overview of current service provision

The CS correctly stated that there are no NICE guidelines for GCA; however, the British Society for Rheumatology (BSR) has developed clinical practice guidelines to advise the diagnosis and management of GCA.[15] The BSR recommends immediate initiation of high-dose glucocorticoid (GC) treatment after suspicion of GCA to prevent sudden vision loss and other ischaemic complications. Diagnosis of GCA should be done using temporal artery biopsy, signs and symptoms of GCA and elevated CRP or ESR levels.

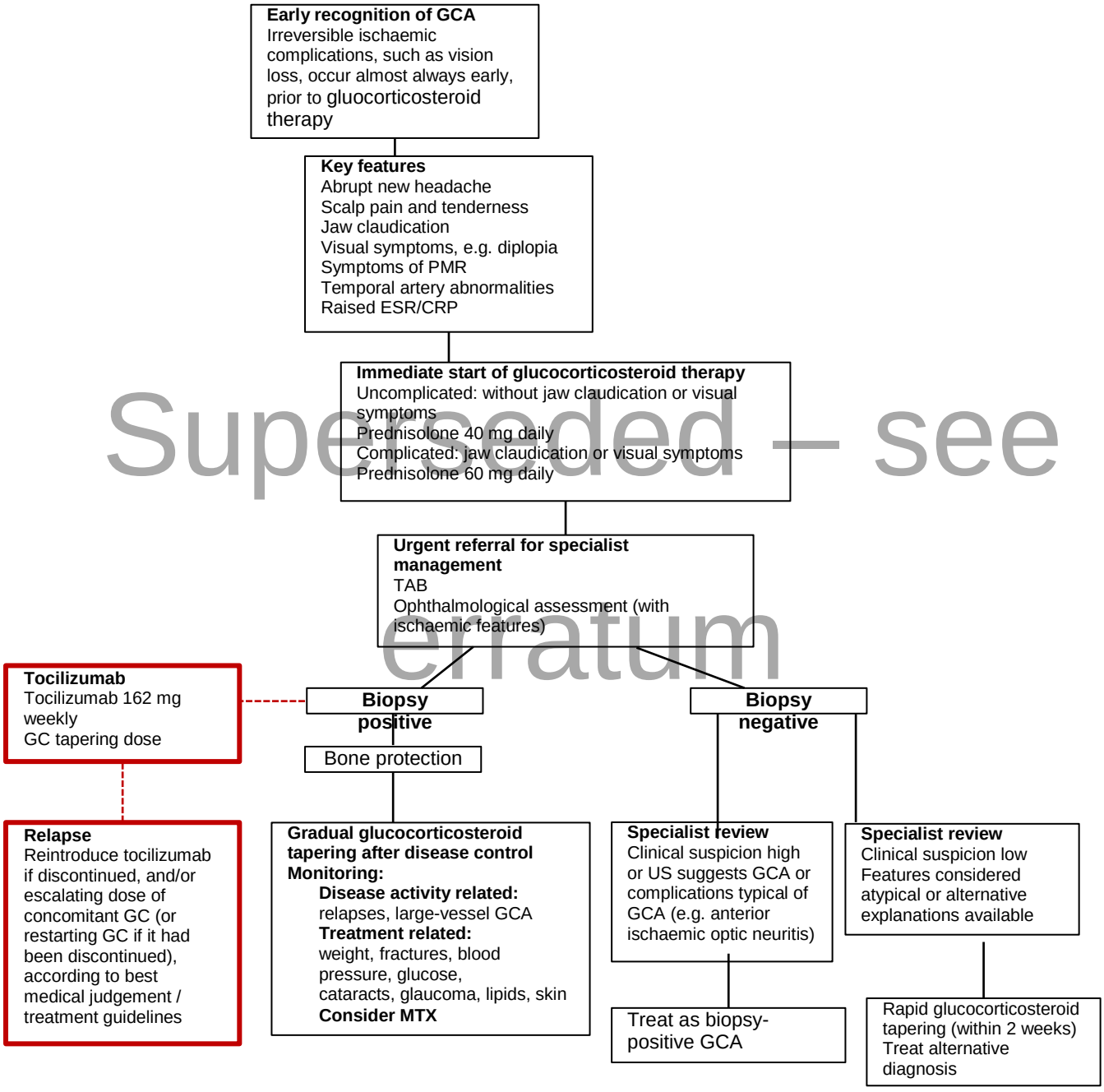
The CS reports that current treatment mainly consists of high dose GC (usually prednisone – the ERG notes that in the UK this is usually prednisolone) followed by long-term steroid tapering.[9, 16] Complicated GCA (evolving vision loss or established vision loss) is treated with an initial dose of 60mg or above, whereas uncomplicated GCA (No jaw or tongue claudication or visual symptoms) is treated with 40-60mg. Once signs and symptoms of GCA are absent patients are slowly tapered off GC. The ERG notes that the BSR recommends a GC tapering regimen which adds up to a minimum of 52 weeks and a cumulative GC dose between 3.6g and 7.4g over approximately 1 - 1.5 years, in

those patients who do not experience a relapse or flare.[16] However, if a patient relapses or flares the GC dose needs to be increased and then tapered accordingly, which can increase the duration of treatment and the cumulative GC dose substantially. The CS states that at least 50% of GCA patients are reported to relapse during GC tapering[17, 18] but also states that the majority of relapses are associated with rapid tapering.[15] However, the ERG notes that patients with GCA rarely relapse while receiving more than 20mg of daily GC; the majority of relapses occur when patients GC dose is tapered to below 10mg/day.[19] Patients receiving a high cumulative dose of GC often experience GC-related adverse effects (AEs) due to the toxicity associated with long term steroid use. The CS stated that approximately 86% of GCA patients experience GC-related AEs after 10 years of follow up.[17] These patients are at an increased risk of developing diabetes, osteoporosis, fractures and serious infections compared to patients receiving a lower dose of GC.[18]

Other immunosuppressive drugs have been investigated and considered as alternatives to GC or as GC sparing drugs; however none have been shown to be effective at inducing and maintaining remission once GC treatment has been discontinued.[20-23] Methotrexate which is an immunosuppressant used in clinical practice has limited and insufficient evidence to support its use in place of GC treatment.[24, 25] Clinical advice to the ERG confirmed that methotrexate is used in clinical practice but only alongside GC treatment, and only because the options for steroid sparing are so limited: there is no good evidence to support the use of methotrexate and it is often poorly tolerated in patients with GCA.

The company's overview of current service provision is generally appropriate and relevant to the decision problem; however, the treatment pathway was not explained clearly. The typical treatment pathway for GCA patients, with the anticipated place of tocilizumab within the pathway, is presented in Figure 1 but suggests that urgent referral for specialist management only happens if urgent GC therapy doesn't work. However, all patients suspected to have GCA receive urgent GC treatment which usually controls the symptoms. The patient's GC treatment is then tapered. Unfortunately, tapering GC can lead to relapse and return of symptoms, and continued treatment with GC is associated with GC side effects and GC dependence. Therefore, the CS states correctly that an effective non-GC therapy that was steroid sparing would be valuable in the treatment of GCA. The CS is proposing that tocilizumab along with a GC tapering dose is introduced after initial treatment with GC. The CS suggests that tocilizumab would reduce the cumulative GC dose received by patients and therefore reduce the GC-related AEs. This may be achieved by lowering the relapse rate and increasing the remission period but also by having a shorter GC tapering regimen alongside tocilizumab.

Figure 1 Typical treatment pathway for advanced/metastatic breast cancer (CS Figure 1 Page 24)



3 Critique of company's definition of decision problem

3.1 Population

The CS described the relevant population as "Adults with Giant Cell Arteritis" This population matched that specified in the NICE scope.

The clinical effectiveness evidence presented is primarily from patients with GCA from the GiACTA randomised controlled trial (RCT). The trial population included adults over 50 years old who had either new-onset GCA or relapsing GCA and only included patients with active GCA disease within 6 weeks of baseline visit. The ERG clinical advisor stated that the GiACTA trial population is generally applicable to patients seen in NHS practice, with the possible exception of the proportion of patients with large vessel GCA. This is because around 40% of patients in GiACTA were eligible primarily on the basis of large-vessel imaging whereas, in the UK around 95% of patients with GCA present with cranial features and relatively few are diagnosed on the basis of large-vessel imaging. However, this difference may relate in part to differences in the availability of vascular imaging in the UK versus countries where services operate on a fee-for-service model. Furthermore, the ERG noted that the mean age of patients in the GiACTA trial was 69 years old, which is lower than the mean age of GCA patients in the UK CPRD data source (73 years). Therefore, the population in the GiACTA trial is not wholly representative of the UK GCA population.

The CS also included one phase II, randomised, double-blind, placebo-controlled trial as supporting evidence. Study NCT01450137 included thirty adult patients with new-onset or relapsing GCA who were randomised to receive GCs and either tocilizumab (20 patients) or placebo (10 patients).

3.2 Intervention

The intervention presented in the CS was tocilizumab, which matches that specified in the NICE scope. The recommended posology is 162 mg of subcutaneous tocilizumab once every week in combination with a tapering course of GC. Tocilizumab can be used alone following discontinuation of GC but is not used as monotherapy for the treatment of acute relapses.

Tocilizumab is currently awaiting marketing authorisation, which is expected in September 2017. The Committee for Medicinal Products for Human Use (CHMP) Positive Opinion was granted on 20 July 2017 for subcutaneous tocilizumab for the "treatment of GCA in adult patients". The FDA approved tocilizumab subcutaneous injection for the treatment of GCA on 23 May 2017.[26, 27]

The GiACTA trial uses the 162 mg subcutaneous dose of tocilizumab as per the licence. In the trial there were two tocilizumab arms: once a week (QW) dosing and once every other week (Q2W) dosing; only the once a week dosing is licensed and therefore, this report will present tocilizumab

results for this dose only. The GC taper used alongside tocilizumab lasts 26 weeks. The ERG notes that this is much shorter than the length of taper used in clinical practice. The tapering regimen recommended by BSR adds up to a minimum of 52 weeks.[15] Clinical advice to the ERG indicated that, in practice, clinicians will aim to achieve this 52 week taper, but a sizeable proportion will flare, and the treatment/taper starts again. The average length of GC treatment is estimated at 2 years. However, clinical advice to the ERG is that in combination with tocilizumab clinicians will seek to taper GC more rapidly than 52 weeks, and quite possibly aim for a 26 week taper, in order to try to benefit from the GC sparing potential of tocilizumab.

The Phase II study (NCT01450137) tocilizumab was delivered by intravenous infusion: 8mg/kg every 4 weeks. This trial is therefore, not directly relevant to the NICE scope.

3.3 Comparators

The comparator presented by the CS was established clinical management without tocilizumab. The comparator used in the GiACTA trial was placebo with either a short (26 weeks) or long (52 weeks) prednisone taper regimen according to a defined schedule.

This matched the NICE scope. The CS clarified that prednisone was used as it is the mainstay of treatment for people with GCA, both in newly diagnosed (new-onset) and in relapsed/refractory GCA. Clinical advice to the ERG confirmed that in the NHS prednisolone is used rather than prednisone. Furthermore, a study of data from the UK-based Clinical Practice Research Datalink (CPRD) found that 99.7% of patients in the UK received prednisolone.[7]

The CS also confirmed in their response to the PFC that prednisolone is recommended in the GCA guidelines. However, the ERG notes that prednisolone and prednisone are highly comparable drugs, prednisone being the metabolic precursor of prednisolone.

The GiACTA trial used two different placebo controls: one with a 26 week GC taper and one with a 52 week taper. Therefore, based on the discussion of UK practice in Section 3.2 the placebo + 52 week GC taper is the more relevant to UK clinical practice, albeit still a little shorter than typically seen in practice.

The BSR also stated that immunosuppressant's such as methotrexate could be used as steroid-sparing agents when combined with GC. However, the CS did not include methotrexate as a comparator, stating that evidence for methotrexate as treatment of GCA is inconsistent. Clinical advice to the ERG confirmed that methotrexate is not effective in treating GCA and is poorly tolerated in older populations. Therefore, in practice it is mainly used as a co-treatment rather than a comparator.

Similarly, the phase II randomised, placebo-controlled study compared tocilizumab to a placebo comparator with a GC taper in both treatment arms.

3.4 Outcomes

The outcomes specified in the CS Decision Problem were:

- Disease remission
- Time to relapse after disease remission
- GC exposure
- Adverse effects of treatment
- Health-related quality of life

These essentially matched the outcomes listed in the NICE scope. However, morbidity (including vision loss) was listed in the NICE scope as an outcome but vision loss was not reported as a separate outcome in the trials but is included as a complication in the economic model. The risk of vision loss is minimised by high dose GC treatment prior to baseline and by escape GC therapy throughout the trial.

The primary outcome of the GiACTA trial was, 'Proportion of patients in sustained remission at Week 52 following induction and adherence to the protocol-defined GC taper regimen'. To meet adherence to the protocol-defined GC taper regimen patients had to be GC free by week 26 (or week 52 according to treatment arm). Remission had to start at week 12: patients not in remission at week 12 were counted as non-responders. The CS stated in their clarification response that week 12 was chosen as the start of remission due to a requirement for a 40 week period of flare-free remission from week 12 through to week 52. The CS stated that, if met, this would provide compelling evidence of the therapeutic benefit of tocilizumab. However, in practice a patient who does not achieve remission by week 12 would not be considered a treatment failure by their physician.

3.5 Other relevant factors

The CS stated that no equality issues related to the use of EP have been identified or are foreseen.

4 Clinical Effectiveness

4.1 Critique of the methods of review(s)

4.1.1 Searches

The ERG considers the literature searches to be generally appropriate and likely to have captured all the relevant records, but has a number of comments as follows.

Reporting

The databases used for the effectiveness review are reported as being MEDLINE and Embase (using the embase.com interface), MEDLINE in Process (using PubMED interface) to identify in-process citations and e-pubs, and CENTRAL (using the Cochrane Library). This is reported in the CS Section D1.1.1 Search Strategy.

The search strategies used in each of the 3 databases are fully reproduced in Section D.1.1.3 and the date that they were conducted is given. The numbers of records retrieved matches the number given in the PRISMA diagram provided on page 43.

Additional searches of conference websites were conducted to identify potentially relevant posters and abstracts and the reference lists of identified studies were reviewed.

Searches of the trials registers ClinicalTrials.gov and the WHO ICTRP were also conducted to find ongoing studies although nothing is reported about the search terms used or whether any studies were identified.

Strategy

The strategy used in MEDLINE and Embase consists of three sections combined with the AND search operator i.e., 1) giant cell arteritis 2) drug interventions and 3) RCT study type.

In the MEDLINE In Process search via PubMED the strategy consists of terms for 1) giant cell arteritis 2) drug interventions and 3) terms for publication status. For Cochrane, the search (correctly) consists of subject terms only.

The ERG does not have access to the embase.com interface, but notes that the overall structure of the search strategies used for MEDLINE and Embase seems to be appropriate: there are no errors in how the sets are combined; and neither are there any typographical errors in the search terms used. However, at line 8 of the Embase/MEDLINE search strategy it is not clear which fields are being searched using the 15 search terms that begin with ‘clinical trial’ and ends with ‘placebo*’. It appears that there is missing notation in these lines e.g. /de or: ab,ti

Additionally, the search of PubMED for In Process MEDLINE citations (reported in Table 2) includes 2 MeSH terms at line 1 Giant Cell Arteritis [MeSH] and line 2 [Adrenal Cortex Hormones]. These are entirely redundant as the search is trying to identify records that will not yet have MeSH indexing attached to them.

A search for grey literature is reported (at end of D.1.1.1) "Keyword-based searches using relevant disease, intervention and study design terms in Google and Google Scholar were also conducted" but no information is given about the search terms used and what was identified through these searches.

4.1.2 Inclusion criteria

The inclusion and exclusion criteria used to select studies for inclusion in the systematic review of effectiveness of treatments for GCA are detailed in Table 4 of Appendix D.1.1.2 of the CS. The ERG considers these criteria to be appropriate. The initial criteria specified long list of interventions, but once the NICE scope was finalised so that the appraisal comparator was 'established clinical management without tocilizumab', infliximab, adalimumab, etanercept, abatacept, sirukumab, immunosuppressants, azathioprine, methotrexate, cyclosporin A and other biologics were excluded from the review. Only English language studies were to be included, but this would almost certainly screen out only secondary publications of trials of tocilizumab.

The results of the screening of the results of the literature searches are presented in Section D1.1.6 and D 1/1/7 and excluded studies with reason are listed.

The ERG does not believe any relevant trials of tocilizumab were missed.

4.1.3 Critique of data extraction

The methods of data extraction are reported in CS Section D1.1.4 and are appropriate.

4.1.4 Quality assessment

The quality assessment of the studies identified for inclusion in the systematic review of effectiveness is reported in Appendix Sections D1.1.9 and D 1.3. The assessment considered the following factors relating to quality and risk of bias:

- Was randomisation carried out appropriately?
- Was the concealment of treatment allocation adequate?
- Were groups similar at the outset of the study in terms of prognostic factors?
- Were care providers, participants, and outcome assessors blind to treatment allocation?
- Were there any unexpected imbalances in dropouts between groups?
- Did the authors measure more outcomes than they reported?
- Did the analysis include an intention-to-treat analysis?

This assessment appeared appropriate and well conducted. Details and further commentary on the results of this assessment are given in Section 4.2.2.

4.1.5 Evidence synthesis

The CS did not present any evidence synthesis. The CS stated correctly that GiACTA was the only randomised clinical study identified in the SLR to be relevant to the decision problem, therefore a standard meta-analysis was not feasible. Furthermore, it would not be appropriate to attempt pooling of the efficacy data from the Phase III GiACTA study and the Phase II NCT01450137 study because of differences in treatment regimens and study designs.

The ERG notes that as the GiACTA trial compared tocilizumab directly with the only relevant comparator, there was no need to include an indirect comparison with other treatments.

4.2 Critique of trials of the technology of interest, their analysis and interpretation

The CS presented two RCTs of tocilizumab in GCA: the Phase II study (NCT01450137) and a Phase III RCT (GiACTA). GiACTA is the only RCT that provides data on the effectiveness of the licensed dose of tocilizumab in patients with GCA. The Phase II study (NCT01450137) provides only supporting evidence. The GiACTA trial includes both newly diagnosed and relapsing patients with GCA.

The ERG did not identify any other directly relevant trials.

4.2.1 Design of the GiACTA trial

Randomised phase (Part 1)

The GiACTA trial investigated the clinical effectiveness of tocilizumab in 251 adults over 50 years old with new-onset or relapsing giant cell arteritis. The trial was preceded by a 6 week screening phase between patients presenting with GCA flare and the trial baseline. During this pre-trial phase the flare was managed with GC, with the aim of achieving remission at baseline. At baseline patients were randomised to one of four arms, two arms of intervention treatment, which were: 162mg of tocilizumab once a week with a 26 week GC taper (TCZ QW+26) and 162mg of tocilizumab every other week with a 26-week GC taper (TCZ Q2W+26) and two arms of placebo treatment: one with a 26-week GC taper (PBO+26) and one with a 52-week GC taper (PBO+52). There were 100 patients in the TCZ QW+26 arm, 50 patients in the TCZ Q2W+26 arm, 50 patients in the PBO+26 arm and 51 patients in the PBO+52 arm.

The trial was a double-blind, randomised, multicentre, placebo-controlled study. The primary efficacy objective was the proportion of patients in sustained remission at Week 52 following induction and adherence to the protocol-defined GC taper regimen. The secondary endpoints included the time to

GCA flare after disease remission, patient reported outcomes (PROs), and health related quality of life (HRQL). A summary of the methods of the GiACTA trial is presented in the CS Table 6 Page 29).

The ERG has the following comments about the design of the GiACTA trial. The CS is unclear about when GC tapering starts in participants. It is the ERG's understanding that patients were not all in remission at baseline (this was confirmed in the company's clarification response), but even so, all patients had to start the tapering protocol. The imposition of a tapering of the GC dose on patients not in remission and who are otherwise receiving placebo seems to be a bias against the placebo arm. This potential difference in treatment arms was not stratified for or accounted for in the analysis.

The GiACTA trial population includes both new-onset GCA (diagnosed <6 weeks before baseline visit) and relapsing GCA (diagnosed >6 weeks before baseline visit and previous treatment with ≥40 mg/day GC [or equivalent] for ≥2 consecutive weeks at any time) patients. The ERG is uncertain whether these two sets of patients would be treated similarly in clinical practice. Clinical advice to the ERG indicated that patients with new-onset GCA generally have a better outcome from GC treatment than patients with relapsing GCA: patients with relapsing GCA already have the burden of previous GC treatment with its cumulative toxicity, meaning that clinicians may be reluctant to go straight to the highest doses; and after initial response to GC, relapsing patients are then more likely to flare again during tapering, because patients who have flared once are more likely to flare again subsequently. Therefore, tocilizumab may be more beneficial in patients with relapsing GCA who have previously been exposed to GC treatment. However, new-onset patients who have experienced adverse effects from GC or are at high risk of mental health problems would benefit from tocilizumab treatment and lower cumulative doses of GC. The ERG also notes that based on a published article on the baseline characteristics of the GiACTA trial,[28] 17% of the trial patients were refractory to GC therapy, i.e. they had never achieved remission with GC.

The intervention in the GiACTA trial was 162mg of tocilizumab in combination with a tapering course of GC, which matched that specified in the NICE scope. The comparator used in the GiACTA trial was placebo in combination with a tapering course of GC. As stated in Section 2.3 the GC tapering regimens in the placebo arms are shorter than recommended practice. Only the placebo+52 week taper can be considered an appropriate comparator.

The primary outcome of the GiACTA trial was proportion of patients in sustained remission at Week 52 (following induction and adherence to the protocol-defined GC taper to reduce GC dose to zero). The primary outcome comparison was with patients receiving placebo + 26 week GC taper. The secondary comparison of the GiACTA trial was the same outcome (sustained remission at Week 52), but compared with placebo + 52 week GC taper. The ERG note that as the placebo+52 week taper is

the more relevant comparator for the present appraisal, this secondary comparison is the relevant one in terms of sustained remission.

Another secondary outcome was time to first GCA disease flare after disease remission, which is a key outcome in the economic model. Flare was determined by the investigator and defined as the recurrence of signs and symptoms of GCA and/or an erythrocyte sedimentation rate (ESR) ≥ 30mm/h attributable to GCA. Remission was defined as the absence of flare and normalisation of C-reactive protein (CRP < 1mg/dL). Patients were not at risk of flare until after remission had been achieved, however, not all patients were in remission at baseline. The CS provided the time to remission for subjects not in remission at baseline, as requested by the ERG. The median time to remission was much higher in the PBO+52 group (22 days) than the TCZ QW+26 group (8 days). This affects the follow-up time available for the time to first GCA disease flare outcome as patients in the PBO+52 group have a shorter period of time during which they are at risk of flare. Therefore, this may bias the time to first flare outcome in favour of placebo; the time to first flare is clearly longer in the tocilizumab group but this may be a conservative result due to the difference in baseline remission.

Other outcomes were annualised relapse rate, which is the number of flares between the first clinical assessment of GCA and the final clinical assessment prior to entry into Part 2, divided by the time period between; and exposure to GC, which was calculated based on a patients starting GC dose, the taper schedule (26-week or 52-week) and the assumption that a patient continued the taper without error.

Long-term follow-up (non-randomised) phase of GiACTA trial (PART 2)

Part 2 of the GiACTA trial is an open-label extension which includes patients from Part 1 who will be followed for an additional 2 years. This part of the GiACTA trial is currently ongoing; however the CS has presented some preliminary results. All patients from Part 1 were entered into the open label extension Part 2. Patients in remission at Week 52 of Part 1 are taken off tocilizumab treatment when entering Part 2 of the trial but are still followed up for maintenance of remission. Whereas, patients not in remission at Week 52 or patients who flare or relapse in Part 2 of the trial are treated with tocilizumab at the discretion of the investigator. Maintenance of remission, incidence of flare/relapse and treatment of flare is recorded during Part 2 of the trial.

4.2.2 Participant flow in the GiACTA trial

A consort diagram of the patient disposition was presented in the CS appendices (Figure 3 page 52). The ERG considers the diagram provided sufficient information on the flow of participants during the GiACTA trial.

There were 251 patients randomised in the GiACTA trial. Patients were randomised 2:1:1:1; 100 allocated to the TCZ QW+26 arm, 50 allocated to the TCZ Q2W+26 arm, 50 allocated to the PBO+26 arm and 51 allocated to the PBO + 52 arm. Overall, 41 patients were withdrawn from blinded study treatment; 18 withdrew from the TCZ QW+26 arm, 9 withdrew from the TCZ Q2W+26 arm, 9 withdrew from the PBO+26 arm and 5 withdrew from the PBO+52 arm. The most common reasons for withdrawal were adverse events (15 patients) and withdrawal of consent by the subject (10 patients). The number of patients who withdrew due to adverse events in the TCZ QW+26 and TCZ Q2W+26 arms was 9 and 3, respectively. Whereas, the number of patients who withdrew due to an adverse event in the PBO+26 and PBO+52 arms was 3 and 0, respectively. Of the 41 patients withdrawn from blinded study treatment, 34 patients discontinued Part 1 of the study: 15 patients in the TCZ QW+26 arm, 8 patients in the TCZ Q2W+26 arm, 6 patients in the PBO+26 arm and 5 patients in the PBO+52 arm. The ERG requested more information on the 8 patients who withdrew due to lack of efficacy: the trial protocol specified that following lack of efficacy of trial treatment patients were given escape therapy (GC) and retained in the trial. In their clarification response the CS stated that the majority of these 8 patients withdrew despite receiving escape GC therapy because their physician wanted to put them on alternative therapy, which was not permitted per protocol (methotrexate or IV steroids). No deaths were reported during the 52-week GiACTA trial.

There were 88 patients at the time of the Part 1 data cut (11 April 2016) who had reached the Week 100 visit of part 2 of the GiACTA trial, which is still ongoing. The duration of follow-up in Part 2 ranged from 48 to 84 weeks. In Part 2 of the GiACTA trial the number of patients in the TCZ QW+26 and TCZ Q2W+26 arms was 33 and 17, respectively. The number of patients in the PBO+26 and PBO+52 arms was 18 and 20, respectively.

4.2.3 Baseline characteristics of the GiACTA trial

The CS presented baseline characteristics for the GiACTA trial population (Table 7, Page 34 of the CS). The ERG notes that there is some lack of clarity in this presentation of the baseline details. Based on a published report of the baseline details,[28] the ERG notes that both the characteristics at diagnosis/screening and actual baseline (week 0 of the trial) need to be considered.

One important baseline characteristic missing from Table 7 is the number of patients with GC refractory GCA (who make up 17% of the total population): it is not clear if they are well balanced across the treatment groups. Another characteristic is whether the patient was in remission at baseline. The ERG queried this and in their clarification the company provided the numbers of patients in remission at baseline, by treatment group: PBO+26 arm 18 (36%); PBO+52 arm 25 (49%); TCZ QW+26 arm 44 (44%). The CS also didn't include the time since diagnosis at baseline; therefore it is unclear if this is balanced between treatment groups.

The CS stated that the baseline demographics between the treatment groups were comparable. However, the ERG believes that there were some imbalances between the two treatment groups:

- The disease duration (days) at baseline was variable across the trial arms: PBO+26 arm (364.7); PBO+52 arm (255.2); and TCZ QW+26 arm (306.8). As disease duration could be associated with difficult to treat disease, this imbalance would favour tocilizumab in comparison with PBO+26, and favour the PBO+52 arm when it is the comparator. The clinical adviser to the ERG also advised that longer disease duration may be indicative of the more difficult to treat patients.
- More patients in the PBO+52 arm had signs or symptoms of GCA (47.1%) compared to the TCZ QW+26 arm (37%). Clinical advice to the ERG indicated that, as symptoms for GCA are generally symptoms of cranial GCA (though not always), this may favour the placebo arm as it suggests there are more patients with (often easier to treat) cranial GCA in the placebo arm.
- There was a higher mean ESR in the PBO+26 arm (28.8) compared to the TCZ QW+26 arm (18.7). This may favour the TCZ QW+26 arm as it suggests that patients in the placebo arm have higher disease activity, which is not as well controlled as patients in the TCZ QW+26 arm
- A larger proportion of patients were diagnosed by temporal artery biopsy in the PBO+26 arm (72%) compared to the TCZ QW+26 arm (57%). Clinical advice to the ERG suggested that this may favour the placebo arm as it would have fewer patients with large-vessel GCA compared to the TCZ QW+26 arm.
- The ERG asked for details of vision loss at baseline in the PFC, as this was not provided in the CS. The company provided the number of patients who had a range of visual manifestations at baseline, which appear to be relatively balanced between treatment arms. The number of patients with visual impairment at baseline was very low; blurred vision was reported for 6% of patients and unilateral blindness was reported in 1 patient in each arm. Patients were treated with high-dose steroids prior to baseline, so their disease may have been less active when compared with diagnosis.

The ERG concluded that there are many baseline imbalances between the treatment groups. However, overall the differences between the arms generally balance out, with no obvious skew or leaning.

4.2.4 Summary of the quality of the included trial

The CS included a quality assessment of the GiACTA trial in accordance with the NICE-recommended checklist for RCTs (Table 9, Page 37 of the CS). The ERG considers that the trial is of relatively good quality; however there are a few issues that may increase bias (Table 1). The trial was appropriately randomised on a 2:1:1:1 ratio using an Interactive Voice Response System (IVRS) and so the number of patients in each arm was relatively even according to the ratio. Treatment allocation was concealed for the whole trial population due to randomisation being done using IVRS and

randomisation was stratified by baseline GC dose (<30mg or ≥30mg per day). Therefore, the risk of selection bias is very low.

The ERG disagrees with the CS's judgement that the two groups were similar in terms of prognostic factors. As discussed above, there were some prognostic factors which were unbalanced between the four arms in the GiACTA trial: these imbalances may slightly reduce the reliability of the study results.

The ERG confirms that the trial was double-blinded: investigators, patients and sponsor personnel were all blinded to treatment assignment. However, the GC tapering was performed in an open-label fashion up to and inclusive of the daily dose of 20 mg/day, which was then switched to double-blind for dosages below 20 mg through to 0 mg. Furthermore, patients experiencing disease flare or those who were unable to adhere to the GC tapering regimen received open-label escape prednisone therapy at a dose of at least 20 mg/day and proceeded with an investigator-defined prednisone schedule in an open-label fashion. Although, the open label use of GC may be perceived as a weakening of the trial blinding, the level of GC dosing can be considered an outcome. Furthermore, the primary outcome and many of the secondary outcomes were objective and so would not be affected by the open-label GC doses. The health related quality of life outcomes could have been affected by subjective responses of participants, increasing the risk of performance bias.

The ERG agrees that there were marginal imbalances in dropouts between treatment groups. Therefore, the risk of attrition bias is very low. Similarly, the ERG agrees that the trial did not appear to measure more outcomes than those reported. The outcomes listed in the protocol are similar to the ones reported in the CSR; however the CS only reported outcomes which were relevant for modelling cost-effectiveness. Thus, the risk for selective outcome reporting is also low. Furthermore, efficacy analyses according to the intention to treat principle were performed, with standard censoring methods used for missing data.

Statistical analysis

The analysis of the GIACTA trial can be criticised because it did not take into account the difference between new-onset and relapsing patients, nor that between those who were in remission at baseline and those who were not. Randomisation was stratified by baseline prednisone dose only. Whilst there was a significant difference in baseline prednisone dose between new-onset and relapsing patients, this stratification will not account for the other differences between the new-onset and relapsing populations.

As discussed in the publication of the GiACTA trial baseline characteristics,[28]at baseline a higher proportion of new-onset patients had their disease controlled (70.6% vs 46.2%), and a lower

proportion had signs and symptoms of GCA (32.8% vs 44.7%) and symptoms of PMR (7.6% vs 30.3%). The publications also highlights that there are important differences between new-onset and relapsing patients in baseline comorbidities, in particular higher weight and BMI in relapsing patients. However, clinical advice to the ERG notes that these differences are likely to be consequences of prior GC therapy in the relapsing group.

Sub-group analyses by disease status at baseline (new-onset or relapsing) for Sustained remission at week 52, for Time to GCA flare, and cumulative GC dose were reported in the CS Appendix E.

Table 1 Quality assessment and risk of bias assessment

NICE Checklist Item	CS Quality Assessment	ERG NICE Checklist QA	ERG Cochrane QA
Was randomisation carried out appropriately?	Yes	Yes	Low risk
Was the concealment of treatment adequate?	Yes	Yes	Low risk
Were the groups similar at the outset of the study in terms of prognostic factors?	Yes (baseline demographics were comparable)	No (imbalances between arms)	N/A
Were the care providers, participants and outcome assessors blind to treatment allocation?	Yes	Yes, however there were open-label GC doses above 20mg and open label GC escape therapy	Low risk for objective outcomes but high risk for subjective outcomes
Were there any unexpected imbalances in drop-outs between groups?	No	No (very marginal drop out imbalances)	Low risk
Is there any evidence to suggest that the authors measured more outcomes than they reported?	No (only those relevant to CE modelling reported in CS)	No	Low risk
Did the analysis include an intention-to-treat analysis? If so, was this appropriate and were appropriate methods used to account for missing data?	Yes	Yes	N/A

Generalisability of the GiACTA trial to NHS clinical practice

The generalisability of the GiACTA trial to the UK GCA population is generally appropriate, however there are some differences:

- The number of patients from the UK in each arm of the trial was requested in the ERG's points for clarification. The company confirmed that there were only 15 patients from the UK who received the study drug in the GiACTA trial. Of these, 7 and 4 patients were in the TCZ-QW+26 and TCZ-Q2W+26 arms, respectively. This is a very small proportion of patients and therefore, the trial population may not be representative of the UK GCA population.

- The GiACTA trial includes both new-onset and relapsing GCA patients. Clinical advice to the ERG indicated that these two subgroups of patients would be treated differently in practice. New-onset GCA patients are typically easier to treat and can often control their disease using GC treatment within one year. Clinical advice suggested that tocilizumab would preferably be used in relapsing patients and new-onset patients who are at high risk of mental health problems, or pre-existing diabetes or osteoporosis /fragility fracture, or those who experience adverse effects from GC. Therefore, the GiACTA trial population may not be wholly generalizable to the population treated in clinical practice.
- The baseline characteristics of the GiACTA population appear to be fairly representative of the UK GCA population. However, the ERG notes that there is an important difference in the mean age of patients in the GiACTA trial (69.05 years) and that from the UK CPRD data source (73 years). The ERG considered that the age reported in the UK CPRD data source more appropriately reflects the relevant population in England and Wales. Also, overall there were a higher proportion of large vessel GCA patients than cranial GCA patients. Clinical advice to the ERG indicated that, in practice, there would typically be more cranial GCA patients. Therefore, there may be an over-representation of large-vessel GCA patients in the GiACTA trial.
- The trial uses a 26 week GC taper for three of the four treatment groups. This is much shorter than that used in UK clinical practice. Clinical advice to the ERG indicated that, in practice, the average length of GC treatment is just over 2 years. Furthermore, the tapering regimen recommended by BSR adds up to a minimum of 52 weeks.[15] Importantly, several studies have shown that both the initial GC dose and the tapering schedule appear to influence the relapse rate. Higher relapse rates have been reported in the context of clinical trials with adjuvant therapies where GC tapering is more aggressive than in routine clinical practice.[13] Consequently, although the 52-week tapering regimen is consistent with the most rapid tapering regimen recommended in the BSR/BHPR guidelines, uncertainty remains concerning the generalisability of this tapering regimen and the associated relapse rate to a longer GC tapering regimen (18-24 months) more conventionally achieved. In summary, the placebo arm with a 52 week GC taper is most relevant to UK clinical practice.

4.2.5 Summary of results of GiACTA

Disease Remission

The primary endpoint of sustained remission at Week 52 of both tocilizumab groups compared with patients receiving placebo + 26 week GC taper was reported on pages 38-39 of the CS. However, the placebo + 26 week taper is not a relevant comparison for UK clinical practice, as in practice a much longer taper of 52 weeks or more is used.

The NHS relevant comparison between TCZ QW+26 and PBO +52 for sustained remission at Week 52 was reported on pages 39-40 of the CS. The number of participants with sustained remission at Week 52 was significantly higher in the TCZ QW+26 arm (56.0%) compared with the PBO+52 arm (17.6%); the difference in percentage of responders was 38.35 (99% CI 17.89 to 58.81) (p<0.0001) (Table 11, Page 40 of the CS). Induction of remission had to occur within 12 weeks of randomisation to meet the sustained remission endpoint. The ERG requested the numbers and proportions of patients who were not in remission at Week 12, which was not reported in the CS. The company provided the number of patients not in remission at Week 12 and the number of patients not eligible for sustained remission, which was lower for both the tocilizumab arms compared to the placebo arms (Table 2). The patients not in remission (but eligible for sustained remission) are participants who achieved remission before 12 weeks and therefore can still meet the primary endpoint of sustained remission: the ERG calculated these numbers and present them in Table 2 for clarity. As stated earlier, the ERG has some concerns that the chance of a placebo patient, who was not in remission at baseline, achieving remission at week 12 was biased against by the imposition of the GC taper from baseline.

Table 2 Patients in remission status at Week 12 (adapted from Table 2 Company's clarification response)

Week 12, n (%)	PBO QW + 26 Week GC Taper (n=50)	PBO QW + 52 Week GC Taper (n=51)	TCZ QW + 26 Week GC Taper (n=100)	TCZ Q2W + 26 Week GC Taper (n=49)
Not in remission at week 12	7 (14.0)	9 (17.6)	7 (7.0)	6 (12.2)
In remission at or before week 12 (eligible for sustained remission	21 (44%)	25 (49%)	83 (83%)	40 (82%)
Not eligible for sustained remission	29 (58.0)	26 (51.0)	17 (17.0)	9 (18.4)

Time to first GCA flare

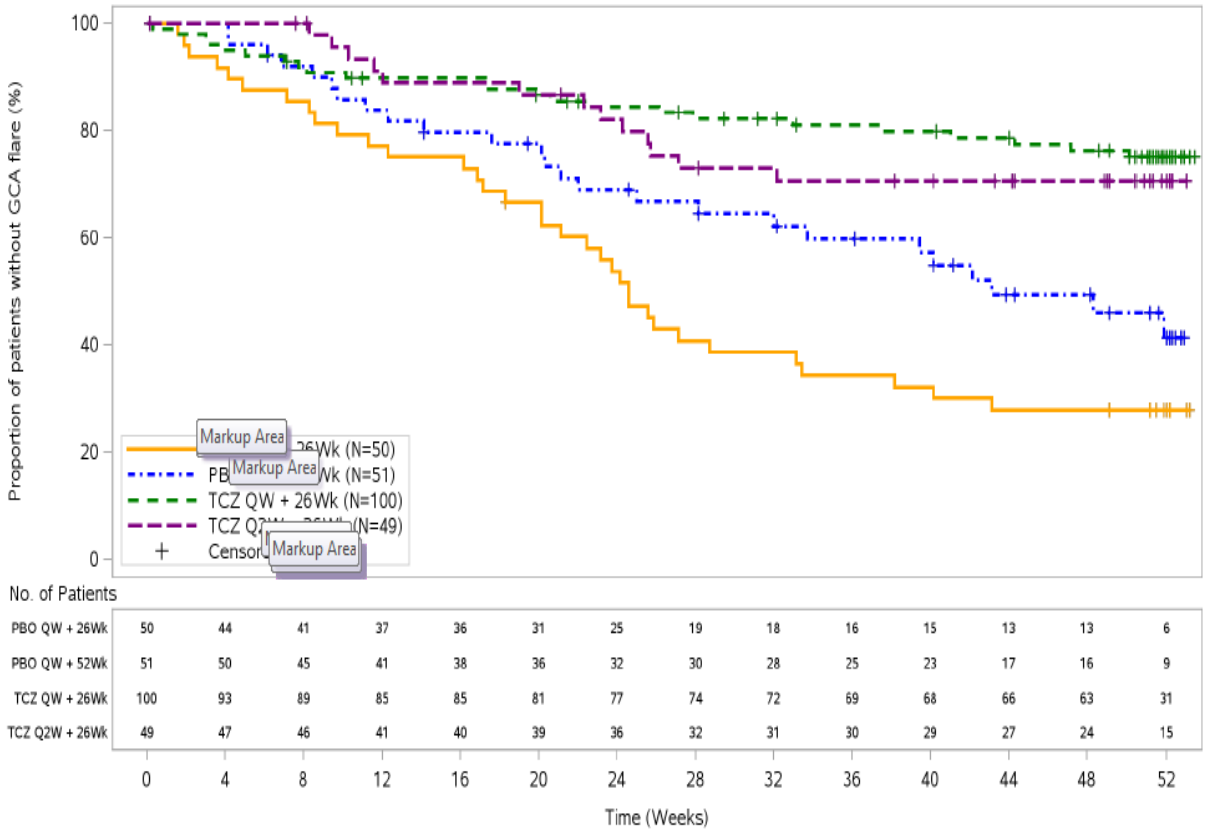
The results for time to first GCA flare are presented in Section B2.6.2 of the CS. The percentage of patients experiencing a flare by Week 52 was less for those in the TCZ-QW+26 arm (23.0%) compared to those in the PBO+52 arm (49%). Tocilizumab treatment significantly increased the time to first flare (not estimable in the TCZ QW+26 arm) compared with PBO+52 arm (295 days %% CI 168 to NE). (Analysis stratified for baseline dose of GC $\leq$ 30mg or > 30 mg/day) HR 0.39 ((%% 0.18 to 0.82) (P=0.0011).

The ERG had some queries about the time to event analysis. The CS states that patients who were never in remission were censored at Day 1. However, the KM plot presented in the CS suggests that was almost never the case. The company response clarified that only 7 patients were censored at Day 1 due to never achieving remission; 2 in each of the PBO+26, TCZ-QW+26 and TCZ-Q2W+26 groups, and 1 in the PBO+52 group. The company also clarified that 'never achieved remission' means the patient never achieved remission throughout the entire study up to Week 52. The Stone et

al. (2017) publication states that patients who never had remission were considered to have had a flare at week 0. The company clarified that the wording was in-part slightly misleading: patients who never achieved remission were censored at Week 0 (Day 1) and so were handled like a withdrawal patient rather than a flare patient. They were censored in this way for all time to flare analysis presentations. However, it may have been more appropriate to treat the 7 patients who were never in remission as flares at Day 1 rather than withdrawals.

The ERG queried the KM plot using time zero as the time when remission was achieved because not all patients who achieved remission did so before or at week 0. The CS provided the proportion of patients who were in remission at baseline (week 0): 64% in the PBO+26 arm, 51% in the PBO+52 arm, 55% in the TCZ QW+26 arm and 59% in the TCZ Q2W+26 arm. The CS also provided a table listing the time to remission for patients not in remission at baseline, by treatment group (See Appendix Table 6). The company clarified that the time to event analysis had used time zero as the time when remission was first achieved post-baseline. The time to flare is calculated as the date of flare minus the date of first remission plus one day. The CS presented an updated KM plot which also accounts for baseline remission, so that patients in remission at baseline will have a time 0 at baseline (Figure 2). The median duration of follow-up whilst at risk of flare was 167.5 days for the PBO+26 arm, 236 days for the PBO+52 arm, 358 days for the TCZ QW+26 arm and 310 days for the TCZ Q2W+26 arm. The revised curves are very similar to those provided in the original CS. The ERG analysed these updated curves as the analyses were not provided by the company. The hazard ratio decreased slightly from the previous KM analysis (HR 0.39 (99% CI 0.18 - 0.82) to HR 0.37 (95% CI 0.2-0.7), similarly showing a statistically significant lower risk of flare in patients in the tocilizumab group compared to the placebo +52 week group ($p<0.001$).

Figure 2 Kaplan-Meier plot of time to first GCA disease flare following clinical remission, by treatment group.



Annualised Relapse Rate

The mean annualised relapse rate for multiple flares observed in each patient are presented in Table 13, Page 43 of the CS. It was 1.30/year in the PBO+52 arm (median: 1) compared with 0.41/year in the TCZ QW+26 arm (median: 0). The median annualised relapse rate was 0 in the TCZ QW+26 treatment group because fewer than 50% of patients had experienced a GCA flare by Week 52.

Exposure to glucocorticoid

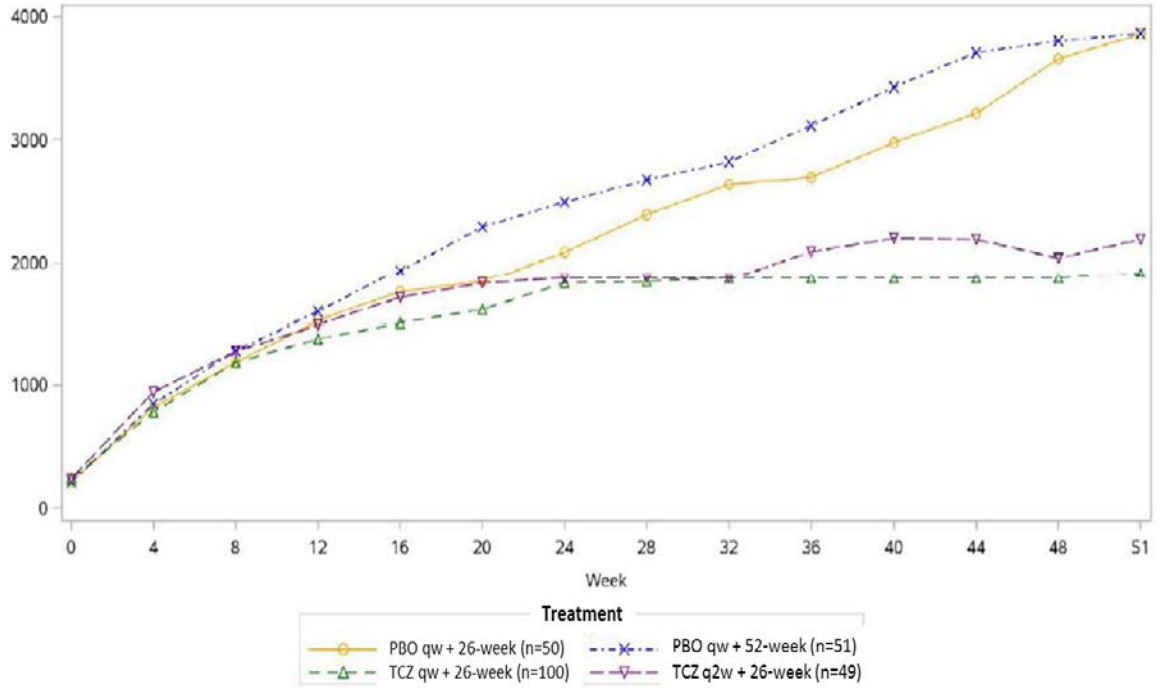
The median cumulative GC dose calculation included the open-label GC taper, blinded GC/placebo as well as escape and commercial GC (for concomitant conditions). It was presented in Table 14 on page 46 of the CS.

There was a statistically significant lower median cumulative GC dose to Week 52 in the TCZ QW+26 group (1862mg) when compared to the PBO +52 group (3817.5mg) ($p<0.0001$). The respective mean values were 2097.84 (SD 1248.45) mg and 4199 (SD 2291.32) mg. The higher cumulative GC dose in the placebo group is probably due to the longer GC taper of 52 weeks rather than 26 weeks and the increased use of escape GC therapy. There was also a notable difference in the initial GC doses taken for new-onset patients and relapsing GCA patients. In newly diagnosed patients 18% had initial GC doses of 60mg/day, whereas only 5% of relapsing patients had initial GC doses of

60mg/day. Relapsing patients who receive lower doses of GC may have a lower chance of achieving remission and thus may be more likely to discontinue GC treatment.

The CS presented a plot of the median cumulative GC dose over time (Figure 3). After Week 22, the curves for the tocilizumab treatment groups start to plateau, whereas the median cumulative GC dose continued to increase in the placebo groups. This may be due to the patients in the tocilizumab groups receiving little additional GC after their GC taper ends and escape patients in the placebo groups receiving increased steroid doses. The proportion of patients receiving GC as escape therapy were lower in the TCZ QW+26 group (23%) compared to the PBO+52 group (55%). However, the difference in median cumulative GC dose between the PBO+52 group and the TCZ QW+26 group can also be attributed to the study design of differing GC taper lengths.

Figure 3 Plot of median cumulative GC dose by visit and treatment group to Week 52 (CS Figure 4)



Health related quality of life

Health related quality of life (HRQL) was measured using four instruments: the Patients Global Assessment (PGA) of disease activity and the SF-36 (a standardised questionnaire of 36 questions) were secondary endpoints; and the FACIT-Fatigue (FACIT-F) score (a self-administered patient questionnaire that consists of 13 statements) and EQ-5D (a generic utility measure used to characterise current health states) were exploratory efficacy endpoints. Information on the completeness of the HRQL questionnaires at each time point was requested in the ERG's points for

clarification. The company provided data for all time points, which appears to be relatively balanced between treatment arms for each HRQL assessment (see Appendix Table 2).

The clinical advisor to the ERG notes that improvements in quality of life over the course of the trial are not necessarily to be expected as patients should have had their symptoms controlled by baseline, though not all were in remission. There was no notable deterioration observed in HRQL in any treatment group, however the tocilizumab groups appeared to score marginally better. Repeated measures methods were used for PGA and SF-36, so all patients were included in the analysis, regardless of their remission status. All treatment groups showed a decline (improvement) from baseline over the 52-week trial for PGA (Table 16, Page 50 of the CS). Whilst, this improvement was more pronounced in the tocilizumab treatment groups the difference between the TCZ QW + 26 group and the PBO+52 group was not statistically significant. The change from baseline to Week 52 of the SF-36 Physical Component score showed a significant improvement in the TCZ QW+26 group, compared to the PBO+52 group, which showed a slight worsening (p-0.0024, Table 17, Page 51 of the CS). Both the TCZ QW+26 group and the PBO+52 groups showed a numerical improvement from baseline in the Mental Component Score; however, there was no significant difference.

In contrast, repeated measure methods were not used for FACIT-F and EQ-5D analyses and patients were censored at flare. There was no substantial deterioration in the EQ-5D scores in any treatment group. The mean changes from baseline were relatively similar between the four groups (Table 15, Page 48 of the CS). Numerically higher mean FACIT-F changes from baseline were observed for both tocilizumab treatment groups versus the placebo groups but no statistical testing was performed. However, the FACIT-F and EQ-5D analyses only provide information on patients in sustained remission and do not reflect the HRQL differences in the entire sample.

Overall, there were only marginal differences between the TCZ QW+26 and PBO +52 groups in HRQL assessments. The only statistically significant differences was seen for and the SF-36 Physical Component Score. Therefore, there is limited evidence to indicate that HRQL improves substantially with tocilizumab compared to placebo. Furthermore, the open label GC escape therapy received by patients experiencing flare may introduce potential bias for the PGA and SF-36 HRQL outcomes; whereas these patients were censored for the FACIT-F and EQ-5D analyses.

Longer term disease control

As stated in Section 4.2.1, Part 2 of the GiACTA trial is an open-label extension which follows patients for an additional 2 years; this part of the GiACTA trial is currently ongoing, with only some preliminary results reported in the CS. Patients in remission at Week 52 of Part 1 are taken off tocilizumab, whereas, patients not in remission at Week 52 or patients who flare or relapse in Part 2 of the trial are treated with tocilizumab at the discretion of the investigator. Maintenance of remission,

incidence of flare/relapse and treatment of flare is recorded during Part 2 of the trial. Data were presented for 88 patients evaluated in Part 2 of the study. Of these, 45 had met the primary endpoint in Part 1(responders) and were therefore followed off treatment in part 2. Of the 35 tocilizumab treated responders in Part 1, 16 patients (46%) flared during Part 2. This indicates that for a sustained treatment benefit, continued treatment with tocilizumab is needed in a substantial proportion of patients.

Subgroup Analyses

The CS reported that subgroup analyses had been performed for 5 pre-defined subgroups: disease onset at baseline (new-onset, relapsing), starting GC dose, previous history of remission, positive imaging with no temporal artery biopsy and no cranial symptoms at diagnosis and GCA diagnosis meeting the ACR criteria. The CS stated that the results of the subgroup analyses were consistent with the results of the overall ITT population; therefore subgroup analyses were not included in the economic model. Sub-group analyses by disease status at baseline (new-onset or relapsing) for Sustained remission at week 52, for Time to GCA flare, and cumulative GC dose were reported in the CS Appendix E.

The ERG notes that a report on the GiACTA trial by Tuckwell et al.[28] divided the trial cohort into newly diagnosed and relapsing patients. The demographic features were similar, but their baseline comorbidities suggested important differences in initial GC dose and remission status at baseline. New-onset patients had higher median starting GC doses than relapsing patients. In newly diagnosed patients 18% had initial GC doses of 60mg/day, whereas only 5% of relapsing patients had initial GC doses of 60mg/day (Table 4). A study by Labarca et al. found that GCA patients treated with an initial oral prednisone dose of >40mg/day achieved earlier prednisone discontinuation than patients treated with <40mg/day.[29] Relapsing patients who tend to receive lower doses of GC may have a lower chance of achieving remission and be more likely to discontinue GC. Therefore, tocilizumab may be more beneficial in relapsing GCA patients than in new-onset patients. Furthermore, 71% of newly diagnosed patients were in remission at baseline, whereas only 46% of relapsing patients were in remission at baseline. This highlights that new-onset and relapsing GCA patients are two subgroups that may require different treatment pathways; this issue is addressed further in Section 5.

Table 3 Median starting GC dose by disease status at baseline (new-onset/relapsing)

	Placebo QW + 52 Week GC Taper (n=51)	**Tocilizumab QW + 26 Week GC Taper (n=100)**
New-onset patients		
n	23	47
Median starting dose	35.0mg	40.0mg
Relapsing patients		
n	28	53
Median starting dose	26.8mg	25.0mg

Sustained remission at week 52 by disease status at baseline (new-onset or relapsing)

The difference in the proportion of patients achieving sustained remission at Week 52 between the TCZ QW+26 group and the PBO+52 group was similar among new-onset and relapsing GCA patients (Table 4). However, the proportion of patients in sustained remission in the PBO+52 group was lower for relapsing patients than for new-onset patients.

Table 4 Sustained remission at Week 52 by disease status at baseline (new-onset/relapsing) (adapted from CS Appendix E Table 10)

	Placebo QW + 52 Week GC Taper (n=51)	sustained remission at Week 52 **Tocilizumab QW + 26 Week GC Taper (n=100)**
New-onset Patients		
n	23	47
Sustained remission	5 (21.7%)	28 (59.6%)
Not sustained remission	18 (78.3%)	19 (40.4%)
Relapsing Patients		
n	28	53
Sustained remission	4 (14.3%)	28 (52.8%)
Not sustained remission	24 (85.7%)	25 (47.2%)

Time to GCA flare by disease status at baseline (new-onset or relapsing)

Kaplan-Meier curves of time to first flare by disease status at baseline (new-onset or relapsing) were presented in the CS Appendix E 1.3 (see Figure 4 and Figure 5 below).

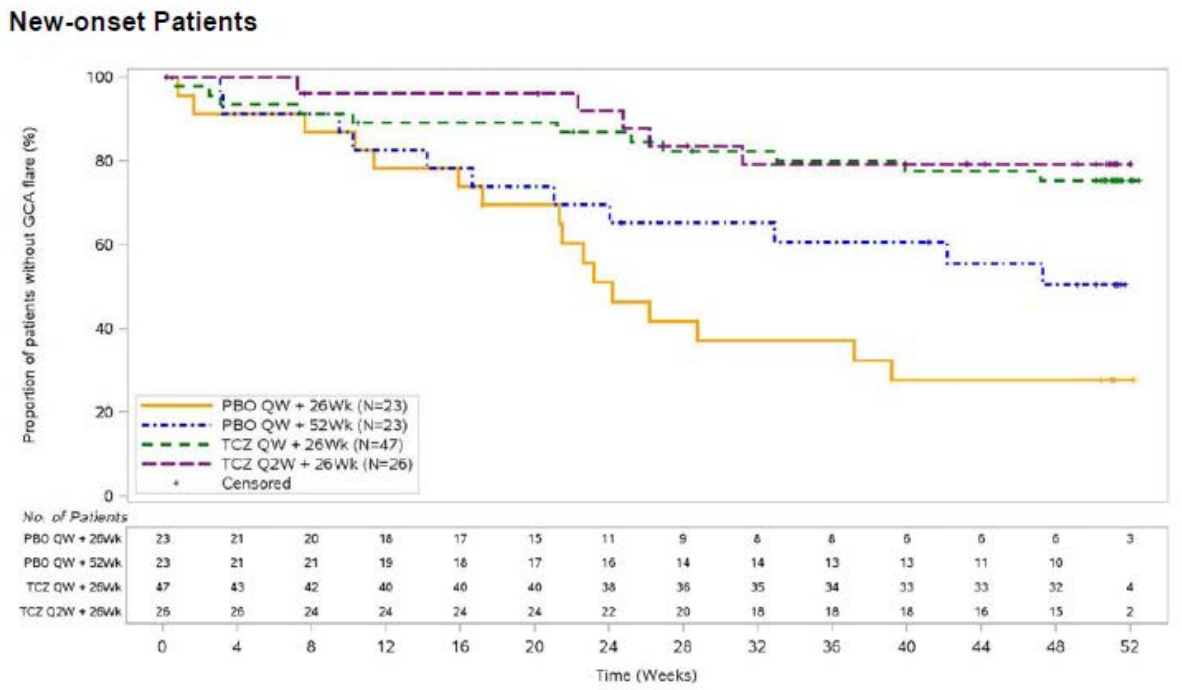


Figure 4 Kaplan-Meier Plot of Time to First Flare in New-onset patients at Baseline

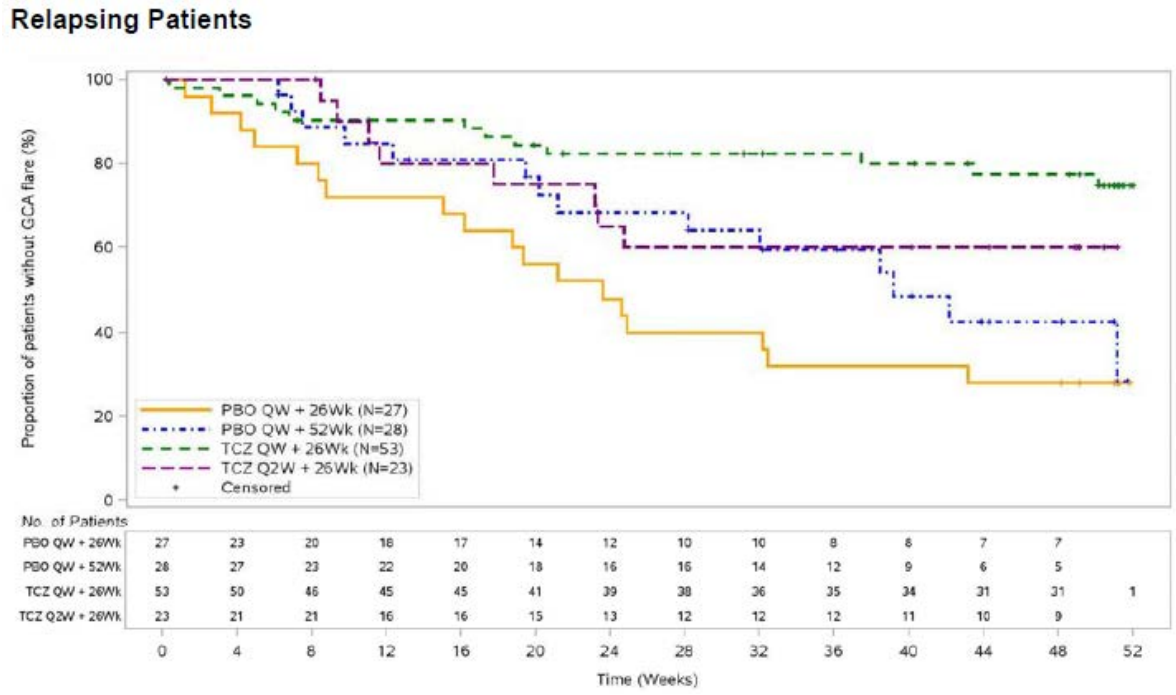


Figure 5 Kaplan-Meier Plot of Time to First GCA Disease Flare in relapsing patients at Baseline

The median time to GCA disease flare in new-onset GCA patients was 169 days in the PBO+26 group and was not calculable for the other three groups due to fewer than 50% of the new-onset patients in

these groups experiencing a flare. In relapsing patients it was 165 days in the PBO+26 group and 274 days in the PBO+52 group but was not calculable in the tocilizumab treatment groups. The CS did not report the hazard ratios for these subgroups and so the ERG performed the analysis. The median time to GCA disease flare in new-onset GCA patients was 169 days in the PBO+26 group and was not calculable for the other three groups due to fewer than 50% of the new-onset patients in these groups experiencing a flare. In relapsing patients it was 165 days in the PBO+26 group and 274 days in the PBO+52 group but was not calculable in the tocilizumab treatment groups. The ERG analysed both subgroups and found that the relative treatment effect was slightly less in the new-onset patients (HR 0.44, 95% CI 0.29 -1.59; (p=0.004)) compared with the relapsing patients (HR 0.33, 95% CI 0.14 – 0.81; (p=0.04)

Cumulative GC dose by disease status at baseline (new-onset or relapsing)

Cumulative GC dose by disease status at baseline (new-onset or relapsing) is presented in the CS Section E1.4. The NHS relevant arms are given in Table 5 below.

Table 5 Cumulative GC dose by disease status at baseline (new-onset or relapsing) (adapted from CS Appendix E 1.4 Table 11)

	PBO QW + 52-week GC Taper n = 51	TCZ QW + 26-week GC Taper n = 100
New-onset		
n	23	47
Mean (SD)	4136.83 (2055.62)	2406.67 (1341.88)
Median	3817.50	1942.00
Range	2017.5–10275.0	630.0–6602.5
95% CI of the Median	2577.5, 4584.5	1822.0, 2519.0
Relapsing		
n	28	53
Mean (SD)	4250.06 (2504.68)	1823.96 (1100.85)
Median	3785.50	1385.00
Range	822.5–10697.5	658.0–5912.0
95% CI of the Median	2222.5, 5372.5	1127.0, 1862.0

The mean differences between cumulative dose in the TCZ QW arm and the PBO+52 arm for these subgroups were not compared formally, but it was numerically higher in the relapsing patients (2426 mg compared with 1730 mg) despite their lower GC dose at baseline (Table 3).

4.2.6 Adverse events of tocilizumab

The CS reported on the adverse events associated with tocilizumab in GCA, which are summarised in Table 21 on page 63 of the CS. The CS presented data on common adverse events, serious adverse events (SAE) and adverse events of special interest (AESI). The total number of patients with at least one AE was similar across all treatment groups; however it was highest in the TCZ-QW group (98.0%) and lowest in the PBO+52-week group (92.2%). The proportion of patients with AEs related to GC was similar in the TCZ-QW (50.0%) and PBO+52-week group (49.0%); similarly, the number of patients with grade 3 AEs was similar in the TCZ-QW group (24%) and the PBO+52-week group (26%).

Fewer patients treated with tocilizumab experienced SAE compared with patients in the PBO+52 group; 15% in the TCZ QW+26 group and 25.5% in the PBO+52 group. None of the SAE were fatal. The proportion of patients with AE leading to withdrawal from blinded treatment was 11.0% in the TCZ QW+26 group, whereas there were no such events in the PBO+52 group. The most common system organ class (SOC) for all-grade AE and Grade 3 AE was 'Infections and Infestations', which was also an AESI based on potential safety concerns associated with tocilizumab. The CS stated there were no marked differences in the overall incidence of patients with infections between the treatment arms. However, the number of patients with 'Infections and Infestations' was notably higher in the TCZ QW+26 group (75.0%) compared with the PBO+52 group (64.7%) (Table 22, Page 64 of the CS). The number of serious infections however, was higher in the PBO+52 group (11.8%) than the TCZ QW+26 group (7.0%). The number of patients with all other AESI was relatively similar between the TCZ QW+26 and PBO+52 groups (Table 23, Page 72 of the CS).

As tocilizumab is given with the intention of being steroid sparing it might be hoped that GC-associated AEs would be lower in the TCZ QW+26 arm. In GiACTA however, the percentage of patients reporting an AE considered related to GC use by the investigator was similar in the TCZ QW+26 (50%) and PBO+52 (49%) groups. More patients in the TCZ QW+26 group had the following GC related AE: infections, general disorders and musculoskeletal and connective tissue disorders when compared to the PBO+52 group (Table 26, Page 75 of the CS). Whereas, more patients in the PBO+52 group had GC related skin and subcutaneous disorders, psychiatric disorders and eye disorders when compared to the TCZ QW+26 group.

Overall, the safety profile of tocilzumab appears to be comparable to the placebo + 52-week GC taper in the GiACTA trial, with a higher number of patients experiencing infections in the TCZ QW+26 group compared with the PBO+52 group.

4.2.7 Phase II NCT01450137 study

In addition to GiACTA a second trial of tocilizumab was identified and presented in the CS: Phase II NCT01450137 study. Details of this trial were presented in Appendix K of the CS. In brief this was a randomised, double-blind, placebo-controlled trial conducted at a single centre: the University Hospital Bern, Switzerland. Similar to GiACTA the population was people aged over 50 years with new-onset or relapsed GCA. The dose and formulation of tocilizumab studies was different to that in GiACTA (licensed). In the Phase II trial tocilizumab was delivered by intravenous infusion: 8mg/kg every 4 weeks. In both trials a tapering dose of prednisone/prednisolone was given in addition to tocilizumab.

The primary endpoint of the Phase II trial was complete remission at week 12 without clinical signs or symptoms of giant cell arteritis, and normal erythrocyte sedimentation rate and C-reactive protein at a prednisolone dose of 0.1 mg/kg per day. Relapse-free survival at week 52 was a secondary endpoint. Other secondary endpoints were time to first relapse after induction of remission, and cumulative dose of prednisolone.

Twenty patients were randomised to tocilizumab and 10 to matching placebo. The baseline characteristics are presented in Table 36. A higher proportion were newly diagnosed (77%) compared to in GiACTA (47%). Three patients discontinued tocilizumab treatment compared to five who discontinued placebo.

The CS does not make it clear what treatment patients in this trial were on immediately prior to baseline; presumably they had been treated with GC to control their symptoms (as in GiACTA). After 12 weeks, 17 (85%) patients in the tocilizumab group and four patients (40%) in the placebo group were still in complete remission, yielding a risk difference of 45% (95% CI 11–79). Adjustment for potential confounders (i.e. age, sex, baseline ESR and CRP) had no major effect on the result. At 52 weeks, 17/20 patients in the tocilizumab group and 2/20 patients in the placebo group were relapse-free. This resulted in an increase of 25 weeks (95% CI 11-39; p=0.0005) of relapse-free survival within the 52 weeks of follow-up of patients in the tocilizumab group. In addition, at Week 52 all 20 tocilizumab -treated patients were in remission, 18 of which had discontinued concomitant GC therapy.

The cumulative weight-adapted GC dose was lower in the tocilizumab patients than in the placebo arm patients, at both weeks 26 and 52: 41 mg/kg vs 66 mg/kg (p=0.0016); and 43 mg/kg vs 110 mg/kg (p=0.0005).

After week 52 tocilizumab treatment was withdrawn and patients were followed for a median time of an additional 12.5 months (range: 3–32 months). Following the last infusion of tocilizumab at Week

52, more than half of the patients (11/20) experienced GCA relapse within a median time of 5 months (range: 2–14).

Thus, the results of this Phase II trial provide supporting evidence for tocilizumab in GCA in terms of greater efficacy and GC sparing, but indicate that in many patients therapy with tocilizumab beyond 52 weeks (maybe chronic therapy) may be necessary.

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

Not applicable

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

Not applicable

4.5 Additional work on clinical effectiveness undertaken by the ERG

Not applicable

4.6 Conclusions of the clinical effectiveness section

The company conducted a systematic literature review and found one relevant RCT which presented clinical data on the effectiveness of tocilizumab. The GiACTA trial was a phase III, randomised, double blind, multicentre, placebo-controlled clinical trial, which was the only directly relevant trial to test the efficacy of tocilizumab. Patients were randomised in a 2:1:1:1 ratio to 162 mg of tocilizumab + 26 week GC taper (TCZ QW +26), 162mg of tocilizumab every other week + 26 week GC taper, placebo + 26 week GC taper or placebo + 52 week GC taper (PBO+52). Only the once a week dosing of tocilizumab is licensed, and therefore, this report presents tocilizumab results for this dose only. Furthermore, the 52-week tapering regimen is consistent with the most rapid tapering regimen recommended in the BSR/BHPR guidelines, and therefore, only the placebo+52 week taper can be considered an appropriate comparator as it is most relevant to clinical practice.

The GiACTA trial was a large, relatively good quality, double-blinded, RCT. However, there were some prognostic factors which were unbalanced between the four arms in the GiACTA trial: these imbalances may slightly reduce the reliability of the study results.

The generalisability of the GiACTA trial to the UK GCA population is generally appropriate, however there are some differences:

- The number of patients from the UK in TCZ-QW+26 the arm of the trial was only 7.
- The GiACTA trial includes both new-onset and relapsing GCA patients. Clinical advice to the ERG indicated that these two subgroups of patients would be treated differently in practice. The

analysis of the GIACTA trial can be criticised because it did not take into account the difference between new-onset and relapsing patients, nor that between those who were in remission at baseline and those who were not. Randomisation was stratified by baseline prednisone dose only. Whilst there was a significant difference in baseline prednisone dose between new-onset and relapsing patients, this stratification will not account for the other differences between the new-onset and relapsing populations. Sub-group analyses by disease status at baseline (new-onset or relapsing) for Sustained remission at week 52, for Time to GCA flare, and cumulative GC dose were reported in the CS.

- The baseline characteristics of the GiACTA population appear to be fairly representative of the UK GCA population. However, the ERG notes that there is a difference in the mean age of patients in the GiACTA trial (69.05 years) and that from the UK CPRD data source (73 years). Also, overall there was a higher ratio of large vessel GCA patients to cranial GCA patients than would be seen in NHS practice.

- The trial uses a 26 week GC taper for three of the four treatment groups. The tapering regimen recommended by BSR adds up to a minimum of 52 weeks.[15] hence, the placebo arm with a 52 week GC taper is most relevant to UK clinical practice. The 26 week taper used with tocilizumab is likely to be attempted in clinical practice, with the aim of reducing the GC load.
- Although the trial included four treatment arms the only comparison relevant to NHS practice is that between TCZ+26 and PBO+52

The number of participants with sustained remission at Week 52 was significantly higher in the TCZ QW+26 arm (56.0%) compared with the PBO+52 arm (17.6%) ($p<0.0001$). Induction of remission had to occur within 12 weeks of randomisation to meet the sustained remission endpoint. However, not all patients were in remission at baseline; 49% in the PBO+52 arm and 45% in the TCZ QW+26 arm. Therefore, the ERG has concerns that achieving remission at week 12 was biased against by the imposition of the GC taper from baseline for patients in the placebo group who were not in remission at baseline. Tocilizumab treatment significantly increased the time to first flare (not estimable in the TCZ QW+26 arm) compared with the PBO+52 arm (295 days %% CI 168 to NE) (HR 0.39 ((%% 0.18 to 0.82) (P=0.0011). Not all patients being in remission at baseline may also bias the time to first flare outcome in favour of placebo.

There was a statistically significant lower median cumulative GC dose to Week 52 in the TCZ QW+26 group (1862mg) when compared to the PBO +52 group (3817.5mg) ($p<0.0001$).

The CS reported that subgroup analyses had been performed for 5 pre-defined subgroups and stated that the results of the subgroup analyses were consistent with the results of the overall ITT population; therefore subgroup analyses were not included in the economic model. However, the ERG believes

that the subgroup analyses of new-onset and relapsing patients should have been a main result. Their baseline comorbidities suggested important differences in initial GC dose and remission status at baseline, highlighting that new-onset and relapsing GCA patients are two subgroups that may require different treatment pathways. The ERG analysed KM plots provided in the CS and found that the treatment effect of tocilizumab relative to placebo was slighter greater in relapsing patients than in new-onset patients when compared to the placebo+52 week group.

The GiACTA trial has an ongoing Part 2, which is an open-label extension including patients from Part 1 who will be followed for an additional 2 years. Preliminary results from Part 2 indicated that for a sustained treatment benefit, continued treatment with tocilizumab is needed in a substantial proportion of patients. Therefore, further reliable research is needed to determine the long term effectiveness of tocilizumab in maintaining remission in patients with GCA.

The safety profile of tocilizumab appears to be comparable to the placebo + 52-week GC taper in the GiACTA trial, with a higher number of patients experiencing infections in the TCZ QW+26 group compared with the PBO+52 group.

5 Cost Effectiveness

This section focuses on the economic evidence submitted by the company and additional information provided in response to the points for clarification. The submission was subject to a critical review on the basis of the company's report and by direct examination of the economic model. The critical appraisal was conducted with the aid of a checklist to assess quality and a narrative review to highlight key assumptions and areas of uncertainty. Section 6 presents additional analyses and scenarios undertaken by the ERG to further address remaining uncertainties.

The company's economic submission included:

- A description of each systematic review conducted to identify published evidence on cost-effectiveness, HRQoL/utilities and resource usage/costs (CS, Sections B.3.1, B.3.4.1 and B.3.5.2), with further details presented in separate appendices (CS, Appendices G, H and I).

- A report on the economic evaluation conducted by the company. The report included: a description of the patient population (CS, B.3.2.1); the model structure (CS, Section B.3.2.2); the clinical parameters used in the economic model (CS, Section B.3.3); the measurement and valuation of health effects and quality-of-life data used in the cost-effectiveness analysis (CS, Section B.3.4), cost and healthcare resource use (CS, Section B.3.5); a summary of the inputs and assumptions used in the model (CS, Section B.3.6); the base-case deterministic cost-effectiveness results (CS, Section B.3.7.1); probabilistic and univariate sensitivity analyses (CS, Section B.3.8.1 and 3.8.2); scenario analysis (CS, Section 3.8.3); the methods of validation (CS, Section 3.10); and the final interpretation and conclusion of the economic evidence (CS, Section B.3.11).

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

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

- A descriptive reply alongside additional data and analyses requested by the ERG.

- An updated Excel-based model including corrections to programming errors, alternative assumptions and additional subgroup analyses based on the ERG's points for clarification.

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

5.1.1 Searches

The electronic databases MEDLINE, MEDLINE In-Process, EMBASE, EconLit, and the Cochrane Library's National Health Service Economic Evaluations Database (NHS EED) were searched via the OVID platform on the 8th of May 2017. The search strategies used for each database were reported in

Appendix G1.3 of the CS. The electronic searches were supplemented with an additional bibliographic review and searches of various disease-specific and HTA congresses and websites.

The structure of the search strategies for MEDLINE, EMBASE and the Cochrane Library were appropriate. Disease terms for GCA were combined with study design terms (e.g. cost-effectiveness, cost-utility) and or other relevant cost and resource utilisation terms.

5.1.2 Inclusion/exclusion criteria used for study selection

The inclusion/exclusion criteria are reported in Table 13 (Appendix G1.3) of the CS. Studies of adult patients (aged 18 years and above) receiving: tocilizumab; any approved or investigational therapy; or established clinical management (including corticosteroids, aspirin and immunosuppresants) were included in the review. Articles were independently assessed by two reviewers against each eligibility criteria and uncertainty regarding the inclusion of studies was checked and judged by a third reviewer.

5.1.3 Studies included and excluded in the cost effectiveness review

A total of 314 potentially relevant articles were identified by the electronic searches and an additional two publications by the supplementary searches. 311 of these articles were subsequently excluded at the primary screening stage. The remaining 5 studies were assessed in full. Only one of these articles was included in the final review.

The single included study was based on a congress abstract and poster.[30] Orfanos *et al*. assessed the lifetime costs and consequences of two tocilizumab doses (TCZ QW and TCZ Q2W) in combination with a 26 week prednisone taper regimen compared to a 52 week prednisone taper regimen alone. The study was undertaken from a UK NHS perspective and used a semi-Markov model. The model used GiACTA trial data to estimate the impact of tocilizumab on disease control (e.g. time in remission and number of flares) and real world data from the US Market Scan Database to estimate the effect of steroid sparing. The real world data was used to quantify the relationship between cumulative prednisone dose and the risk of steroid related adverse events in GCA patients.

Although the study was formally stated to be a cost-effectiveness analysis, the study design is more appropriately defined as a cost-consequence analysis since a range of separate outcomes (or consequences) are presented and there is no attempt to combine these into a single outcome measure (e.g. LYG or QALY) .

The study reported that both doses of tocilizumab used with a 26-week prednisone tapering regimen appeared cost saving compared to a 52-week prednisone tapering regimen alone. Mean per-patient lifetime cost savings ranged between £3,255 (TCZ Q2W+26) and £3,530 (TCZ QW+26). Both tocilizumab strategies were also reported to improve GCA control (i.e. fewer relapses/flares, longer duration of sustained remission and less GCA associated adverse events) with a lower incidence of

steroid related adverse events compared to prednisolone alone. Based on these findings the authors conclude that tocilizumab is cost-effective.

The model presented by Orfanos *et al.* shares an identical structure and many common inputs and assumptions to the company model reported in the CS. The main differences between the previously published model and the company model are: (i) the company model only includes the weekly (TCZ QW+26) dose of tocilizumab based on the CHMP positive opinion; (ii) the company model uses UK specific data from the Clinical Practice Research Datalink (CPRD) to estimate the impact of a steroid sparing effect of tocilizumab; (iii) the study by Orfanos *et al.* appears to exclude the additional acquisition and monitoring costs for the tocilizumab strategies; (iv) the company model combines the separate outcomes into a single QALY measure.

A full critique of Orfanos *et al.* is not feasible given the limited details reported in the abstract and poster. However, the ERG considers that the apparent exclusion of the additional acquisition and monitoring costs from this study to be an important limitation and conclusions regarding the cost-effectiveness of tocilizumab cannot be appropriately drawn.

5.1.4 Conclusions of the cost effectiveness review

The company's search identified a single published cost-effectiveness study of TCZ QW+26 and TCZ Q2W+26 for the treatment of GCA. Given the close relationship between the previously published study and the current submission, the ERG considers that the cost-effectiveness analysis reported in the submission to be the most relevant source of evidence to inform the decision problem.

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

An overview of the company's economic evaluation is presented in Table 6. The results of the checklist used to assess the quality of the submission are reported in Appendix table 3.

Table 6: Summary of the company's economic evaluation

	Approach	Source / Justification	Location in CS
Model	Semi-Markov model with weekly cycles. No half cycle correction was performed due to the short cycle length.	The conceptualisation of the model was stated to have been informed by the disease aetiology, trial data, NICE Scientific Advice and expert opinion (clinician and HTA).	B.3.2.2; p90-95
States and events	Seven health states: • On remission – on steroid • On remission – off steroid • On relapse/flare • On remission – on maintenance steroids • GCA-related complications	Separate remission states were used before a first flare and following the first flare to account for different transition probabilities and GC exposure based on GiACTA trial data. GCA-related complications (vision loss and stroke) were	B.3.2.; p90-95

	• Steroid-related AEs • Death	assumed to only occur from the relapse/flare state and transitions were derived from external literature. Steroid-related AEs included fractures and diabetes based on cumulative GC dose and evidence from real world data using CPRD. Death included background mortality (general population, age and gender matches) arising from any state with an adjustment for stroke related mortality attributed to GCA-related complications.	
Comparators	Tocilzumab (TCZ QW – weekly dosing over a 2-year fixed treatment duration) and prednisone (26-week tapering) Prednisone alone (52-week tapering regimen; PBO+52)	TCZ-QW was assumed to be continued over a 2 year fixed treatment period This was justified based on the CHMP Positive Opinion which states that TCZ can be continued beyond 1-year, clinical opinion and the typical duration of conventional treatment for GCA with GCs. The 52-week GC tapering regimen included in the GiACTA trial was considered a relevant comparator and was consistent the most rapid GC tapering regimen recommended in the BSR guidelines. Other immunosuppresants were not formally included as alternative strategies but some usage was assumed based on utilisation within the GiACTA trial.	B.3.2.3; p96-97
Natural History	Transition probabilities from the initial remission state to the first flare for prednisone alone (52-week tapering) were based on an individually fitted parametric model using patient-level data from the ITT population of the GiACTA trial. Transition probabilities from the subsequent remission state to flare were based on a separate Poisson regression.	An exponential distribution was assumed for the time to first flare based on statistical tests, visual inspection and external expert input. A separate Poisson regression was used to estimate the weekly probability of subsequent flare based on a post-hoc analysis of time at risk and events in the subgroup of patients experiencing an initial flare.	Section B3.3; p99-109
Treatment effectiveness	Transition probabilities from the initial remission state to the first flare for TCZ-QW (plus prednisone 26-week tapering) were based on an individually fitted parametric model using patient-level data from the ITT population of the GiACTA trial.	A weibull distribution was assumed for the time to first flare based on statistical tests, visual inspection and external expert input. A separate Poisson regression was used to estimate the weekly probability of subsequent flare based on a post-hoc analysis of	Section B.3.3.2; p100-103

	Transition probabilities from the subsequent remission state to flare were based on a separate Poisson regression.	time at risk and events in the subgroup of patients experiencing an initial flare. The effectiveness of tocilizumab was assumed to be maintained over a lifetime and justified based on early results from open label data.	
Adverse events	The risk of GCA related complications (vision loss, stroke) was derived from external literature and only applied to the flare/relapse state. The risk of GC related complications (diabetes, fracture) was based on cumulative GC dose burden and external evidence from the literature reporting the association between different levels of GC dose and the associated risk of fracture and diabetes.	The risk of GCA related complications was assumed to be related to subsequent relapse/flares. In the absence of these complications arising in the GiACTA trial, estimates were sourced from a separate published economic model comparing alternative diagnostic approaches for GCA. The use of external evidence was justified due these events being rare but associated with significant costs and HRQL implications. Cumulative GC dose for each treatment arm were based on 3 separate estimates to reflect dosing during: (i) the initial remission period (prior to first flare), (ii) during secondary remission (post-initial flare) and (iii) during relapse/flare. Dose estimates were based on data from the GiACTA trial and real world evidence.	Sections B.3.3.5. & B.3.3.6; p106-109
Mortality	Background mortality was assumed to be the same as the general population. An adjustment was made to avoid double counting the mortality attributed to stroke.	Background mortality was derived from standard lifetables and justified based on findings from a systematic review which found no significant differences in mortality for GCA patients.	Sections B.3.3.8. & B.3.3.9; p109
Health-related quality of life	Separate utilities were applied to the remission and relapse/flare states (4-weeks only). Additional utility decrements were applied to GCA and GC related complications.	Utilities for the remission and relapse/flare states were sourced from a mixed effect regression model based on EQ-5D data from GiACTA. Data was combined across the separate arms and justified given the lack of significant difference by treatment arm reported within the trial. The relapse/flare utility was applied for a 4-week duration based on published literature and clinical opinion. Utility decrements for GCA and GC-related complications were sourced from the external literature.	Section B.3.4.5; p115-117

Resource utilisation and costs	The treatments costs of tocilizumab and GC treatment included the acquisition, administration and monitoring costs. Separate heath state costs were applied based on remission status and associated use of steroids (on/off steroids and on maintenance steroids) and flare episodes. Additional costs were also assigned to GCA related complications and GC related AEs.	The treatment costs of tocilizumab and GC were based on published prices. A separate analysis was reported based on the approved PAS for tocilizumab. The cost of conventional GC treatment was based on published prices for prednisone. Health state costs were based on third-party market research undertaken by the company. The costs of GCA related complications and GC related AEs were derived from the external literature.	
Discount rates	3.5% for costs and outcomes	NICE reference case	Section B.3.2.2; p95
Population and Subgroups	The model only considers the overall ITT population.	The overall ITT population was justified as being the most relevant to the decision problem based on the marketing authorisation and NICE scope. Results were not presented for each of the 2 patient subgroups identified within the NICE scope (newly diagnosed and relapsed/refractory). This was justified based on the favourable cost-effectiveness results for the overall population, the lack of difference in efficacy reported between the subgroups and the lack of statistical power. Separate results for these subgroups were subsequently provided and included in the company response to the points for clarification.	Section B.3.9; p141-142
Sensitivity analysis	Univariate and probabilistic sensitivity analysis and scenarios.	NICE reference case	Section B3.8; p131-141

Key: GCA: Giant Cell Arteritis; ITT: Intention To Treat; GC: Glucocorticoids; AE: Adverse Events; Service; NICE: National Institute for Health and Care Excellence

5.2.1 Model structure

The submission is based on a semi-Markov model using a weekly cycle length. The conceptualisation of the model is stated to have been informed by the disease aetiology, trial data, NICE Scientific Advice and expert opinion (clinician and HTA).

The model structure is shown in Figure 6 and includes seven separate health states:

- On remission and on steroid;
- On remission and off steroid;
- On relapse/flare;
- On remission and on maintenance steroids (escape);
- GCA-related complications;
- Steroid-related AEs;
- Death.

The submission states that people with GCA enter the model either on relapse/flare or in the remission state and treatment is then initiated with TCZ QW plus prednisone or prednisone alone. After achieving remission, patients then follow the GiACTA protocol for steroid tapering (26 weeks for TCZ QW and 52 weeks for prednisone alone) and remain in remission until their first flare.

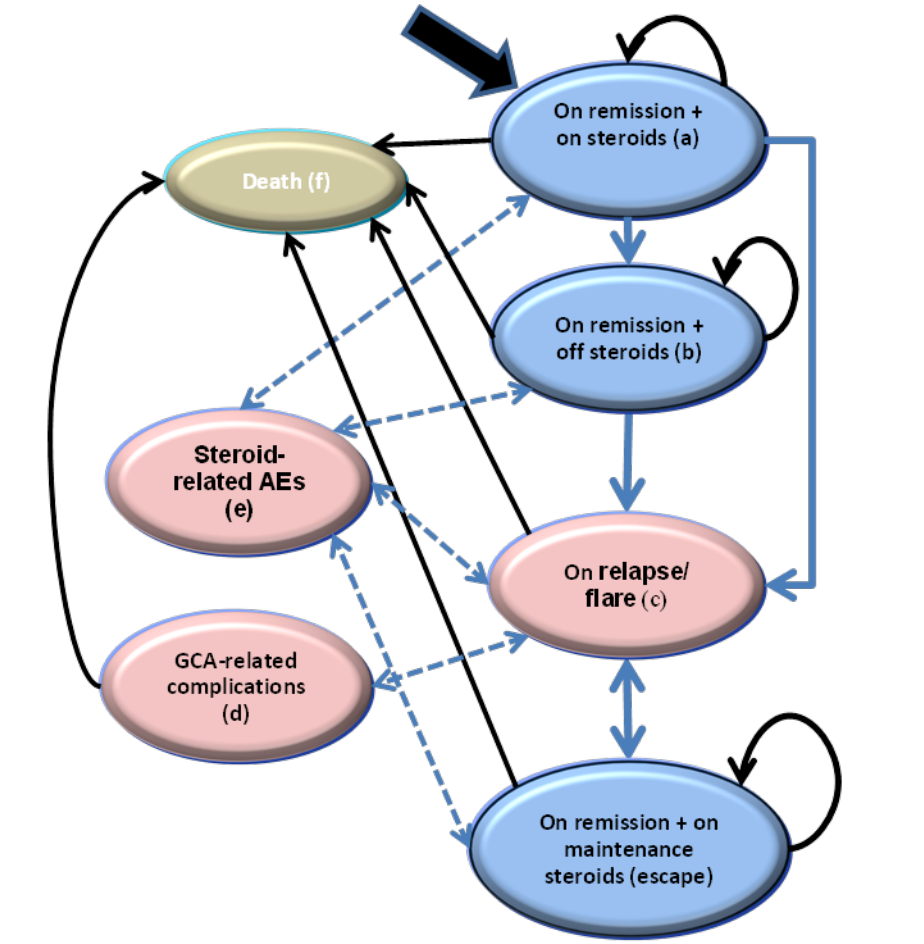
Transitions from the initial remission state are estimated via time-dependent transition probabilities. These probabilities are estimated using parametric survival analysis based on the Kaplan-Meir data from the GiACTA trial on time to first flare. The use of parametric survival analysis allows the probability of an initial flare to be time-dependent and provides a basis for extrapolation beyond the 52-week follow-up of the GiACTA trial.

Following a first flare, patients then transition to a separate remission state – 'On remission and maintenance steroids (escape)'. The separate remission state is used to distinguish the initial remission period from subsequent remission periods. This separation permits different transition probabilities to be assigned within these periods. The probability of further relapse/flare events following a subsequent remission was estimated using a separate Poisson regression based on data from the subgroup of patients following an initial flare from the GiACTA trial. A key assumption of the model is that the probability of a relapse/flare during each subsequent remission is higher than the probability during the initial remission period and is constant with time.

The separate remission and relapse/flare states are used to characterise the natural history of GCA. Separate transition probabilities for TCZ-QW+26 and PBO+52 are used to quantify the impact of the alternative treatments in terms of GCA symptom control (i.e. duration of initial and subsequent remission and number of relapse/flare episodes). Additional states are also incorporated to capture GCA-related complications (visual loss and stroke) and the potential steroid sparing effect of tocilizumab in terms of reducing GC-related AEs (fracture and diabetes).

The four specific GCA complications and GC adverse events selected were based on a wider set of events included in a previous published model and restricted to those which were considered to have the largest impact on HRQoL and costs. The company considered this approach to be conservative as many other GC-related AEs that could be impacted by the GC-sparing effect of tocilizumab were excluded.

Figure 6: Company model structure



a. 26 weeks for patients receiving tocilizumab and prednisone and 52 weeks for patients receiving prednisone alone.
b. Patients who have not yet flared after the end of tapering.
c. Transition probabilities derived from the Kaplan–Meier curves of time to first flare from the GiACTA trial
d. Stroke and vision loss
e. Fractures and diabetes.
f. Background mortality and GCA mortality taken into account

Figure replicated from company submission

The probabilities of GCA-related complications are based on a previously published model and structurally linked to the relapse/flare state. Each time a patient experiences a flare (during the initial or subsequent remission periods) they are assumed to face a risk of experiencing visual loss and/or stroke as a result of the flare. Structurally the model assumes a surrogate relationship between GCA-

related complications and relapse/flare events and that the risks of these complications are modifiable with tocilizumab treatment through a lower risk of relapse/flare.

The probabilities of GC-related AEs were derived from published real world data from CPRD reporting the association between cumulative steroid burden in GCA patients and the rate of fracture and diabetes. Structurally the model assumes that GC-related AEs can be experienced by a patient in any of the remission or flare states. However, the cumulative steroid burden calculations are not directly linked to the individual model states and so the same probability of GC-related AEs is applied to all states during each cycle.

The model assumes no excess mortality risk relative to the general population other than that arising due to one of the GCA complications (stroke). A separate death state is used to capture background (general population) mortality adjusted for stroke-related mortality. The company justified this approach based on a published systematic review which reported no overall increase in long-term mortality for GCA patients.[2]

The ERG considers that the general structure of the model is appropriate and adequately justified by the company. However, the company description of the model structure could have more clearly distinguished between events which are represented using separate and mutually exclusive health states and events which impact the state values or 'rewards' assigned to these states (i.e. cost and HRQoL implications of residing in, or transiting between, the main mutually exclusive health states). Two of the seven health states (steroid-related AEs and GCA-related complications) are not modelled as distinct health states but rather as events which impact the health state values or 'rewards' attributed to other health states and transitions. For example, GCA-related complications are included as events which impact the health state values assigned to a proportion of patients at the point they transition from a remission state to the relapse/flare state. Similarly, GC-related AEs are included as events which impact health state values for a proportion of patients within the remission and relapse/flare states.

In a similar vein, while Figure 6 depicts separate states for the initial remission period (on and off steroids), only a single remission state is actually implemented and the proportion of patients on and off steroids are used to adjust the cost and HRQoL values of the initial remission state.

The ERG's view is that the model is more appropriately described in terms of the following four main mutually exclusive health states:

- On initial remission;
- On relapse/flare;

- On subsequent remission;
- Death;

Other events such as GC-related AEs, GCA-related complications and the proportion of patients on and off steroid treatment during the initial remission period only impact the health state values attributed to these four main states.

The model uses a 1-week cycle length which is justified by the company as being in line with the dosing schedule for TCZ QW and sufficiently short that a half-cycle correction is not required. However, in determining an appropriate cycle length, the frequency of clinical events should also be considered and the cycle length should be short enough that relevant events occur at most once per cycle.[31] While a weekly cycle appears appropriate in the context of the events included in the model (i.e. multiple relapses/flares during a single week does not appear clinically reasonable), the ERG's view is that the use of a single state for the relapse/flare event may not be appropriate in the context of this short cycle length.

Structurally the model only permits patients to reside in the relapse/flare state for a single weekly cycle, whereas the associated health state values are assumed to apply over a longer duration (28 days for the duration of flare disutility and 3 months for the additional resource consequences). As a result, there appears to be an inconsistency between the structural assumptions of the model and the duration of the state values (i.e. HRQoL and costs) assigned to the relapse/flare state. This inconsistency could have been avoided by either retaining a single state for relapse/flare and employing a longer-cycle length or by creating a series of additional (tunnel) states for the flare event (e.g. relapse/flare week 1, relapse/flare week 2 etc.) and retaining the weekly cycle length.

Rather than addressing this inconsistency by structurally changing the model or altering the cycle length, the company applies a series of adjustments within the Excel model itself. These adjustments were performed by initially assigning values which captured the full duration of the HRQoL impact (28 days) and costs (3-months) of the flare/relapse event to the weekly cycle in which the event occurred and then attempting to exclude these patients from the remission state for 4 weeks in the QALY calculations to avoid double counting the same period already captured by the relapse/flare state.

The ERG identified several concerns with the nature of these adjustments as well as a significant programming error. The error was considered to have a potentially important effect on the accuracy and validity of the overall QALY estimates and the associated ICER results. These concerns are summarised below:

- The ERG considers that the adjustments introduce unnecessary programming complexities that could have been avoided by using alternative structural assumptions (e.g. alternative cycle length and/or use of tunnel states).

- An important error was also identified by the ERG in the QALY calculations. Patients who experienced a relapse/flare were only assigned the utility value associated with this state for a single week rather than the full 28-day period stated in the submission. However, these patients were subsequently excluded from the remission state for 4 weeks in the QALY calculation. This means that each time a patient experience a relapse/flare, 3 of the 4 weeks of HRQoL associated with this state are excluded. The impact of this error is likely to significantly under-estimate the QALYs attributed in the model to the relapse/flare state, creating a potential positive bias in favour of tocilizumab given the lower frequency of relapse/flare events assumed for this treatment.

- The adjustment to the QALY calculations in the subsequent remission state avoids one source of double counting. However, the inconsistencies also give rise to another potential source which is not considered. In transitioning patients to the subsequent remission state after only 1 week in the relapse/flare, these patients immediately face the risk of a further relapse/flare. That is, although the duration of a relapse/flare episode is assumed to impact on HRQoL for 4 weeks, the model structure means that patients are at risk of repeat relapse/flare events after 1-week of their event. The ERG was doubtful regarding the clinical plausibility of this.

- Although an adjustment was made to avoid double counting within the QALY calculations, a similar adjustment does not appear to have been undertaken in terms of costs. Hence, patients who experience a relapse/flare appear to be assigned the full 3-month cost during the weekly cycle in which they reside in the relapse/flare state. However, in the following cycle these patients then transition to the subsequent remission state and continue to accrue the weekly costs of this state without any adjustment for the period of time already accounted for by assigning the full 3-month cost estimate following a relapse/flare. Hence, these patients are then assigned an additional 11 weeks of cost in the remission state. This appears to significantly over-estimate the costs attributed in the model as a result of relapse/flare and creates a potential positive bias in favour of tocilizumab given the lower frequency of relapse/flare events.

These concerns were raised with the company as part of the clarification stage and revisions were requested. In their response, the company acknowledged the errors identified by the ERG in the QALY calculations and provided a corrected and updated model and a complete set of revised results. The ERG was satisfied with the corrections but retains the view that a monthly cycle length or tunnel states would have been more appropriate. These structural changes would also have avoided the issue

that patients face the risk of a further relapse/flare after 1 week. However, the ERG does not believe that this issue creates a significant bias and considers the approach sufficient for decision-making purposes.

The company's response also addressed the concerns regarding the lack of a similar adjustment applied to the cost calculations. The company clarified that the costs assigned to the flare/relapse state were considered to represent additional costs that would be incurred on top of the background management costs applied to the remission states. The ERG considers that the implementation in the Excel model is consistent with the company's response. However, the ERG notes that uncertainty remains regarding whether it is appropriate to include these background costs in addition to the 3-month event cost assigned to the relapse/flare state.

The submission states that patients enter the model either on relapse/flare or in the remission (and on steroid) state. However, all patients in the Excel model actually start in the remission (and on steroid) state. The initial transitions (i.e. remaining in remission or experiencing a first relapse/flare event) are informed from the Kaplan-Meier data (ITT population) reported in the GiACTA trial on the time to first flare after clinical remission of GCA. The reason for the apparent discrepancy between the wording of the submission and the implementation in the Excel model is not explained in the submission.

The use of the Kaplan-Meier data within the model raises several issues. Firstly, not all patients in the GiACTA trial had achieved clinical remission at the start of the study and secondly several of these patients never achieved remission during the course of the follow-up. The second issue appears to be captured within the time to first flare Kaplan-Meier data as these patients are treated as an event which occurs at day 1. However, for those patients who were not in remission at the baseline assessment but then subsequently achieved remission, the time period prior to this remission does not inform the Kaplan-Meier data or the model inputs.

These issues were also discussed in the clinical effectiveness review and further clarification and additional Kaplan-Meier data were provided by the company (See Section 4.2.5). The ERG notes that the additional Kaplan-Meier data was not incorporated in the revised model. However, although the period prior to remission is not formally captured in the model, the ERG does not consider that this leads to any significant bias as the evidence does not suggest that this period is longer with tocilizumab and that the approach used may be argued to be conservative.

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

Table 7 summarises the ERG's assessment of whether the company's economic evaluation meets NICE's reference case.

Table 7: NICE reference case and commentary

Attribute	Reference Case	Included in CS	Comment on whether *de novo* evaluation meets requirements of NICE reference case
Comparator(s)	The NICE scope defined the comparator as 'established clinical management'.	Partially	The comparator included in the model was based on the 52-week tapering GC regimen in the GiACTA trial. Although the 52-week tapering period is consistent with the most rapid taper regimen advocated by the BSR/BHPR guidelines, clinicans typically will use a longer tapering regimen in routine clinical practice (18-24 months). Hence, there exists some uncertainty regarding the generalisability of the results from the 52-week tapering regimen to conventional NHS practice. The company's economic evaluation is based on the same GC regimen (prednisone) used within the GiACTA trial. However, prednisolone is more commonly used within the NHS and has a lower acquisition cost than prednisone.
Type of economic evaluation	Cost-effectiveness analysis	Yes	
Perspective - costs	NHS and PSS	Yes	
Perspective - benefits	All health effects on individuals	Yes	
Time horizon	Sufficient to reflect any differences in costs or outcomes between the technologies being compared.	Yes	The economic model is stated to be a lifetime. This is assumed to be 30 years which appears reasonable based on the average age at baseline (69.05 years) and the potential lifelong consequences of complications and adverse events.
Synthesis of evidence on outcomes	Systematic review	Yes	
Outcome measure	QALYs	Yes	
Health states for QALY measurement	Described using a standardised and validated instrument	Yes	Utilities for the remission and relapse/flare states were sourced from a mixed effect regression model based on EQ-5D data from GiACTA. Utility decrements for GCA and GC-related complications were sourced from the external literature.
Benefit valuation	Time Trade Off or Standard Gamble	Yes	
Source of preference data	Representative sample of the public	Yes	
Discount rate	3.5% on costs and health benefits	Yes	
Equity weighting	An additional QALY has the same weight regardless of the other characteristics of the individuals receiving the health benefit	Yes	
Sensitivity analysis	Probabilistic sensitivity analysis	Yes	Probabilistic sensitivity analysis was conducted as well as deterministic sensitivity analyses. Mean increment results for the probabilistic sensitivity analysis were presented as well as graphical results using scatter plots, cost-effectiveness acceptability curves and tornado diagrams.

5.2.3 Population

The economic model was based on the overall ITT population in the GiACTA trial. Separate analyses were not provided in the initial company submission for the two main patient subgroups identified within the NICE scope (newly diagnosed and relapsed/refractory). The company justified their focus on the overall ITT population based on the favourable cost-effectiveness results for the overall population, the lack of difference in efficacy reported between the subgroups and the low statistical power.

The ERG considers the exclusion of these patient subgroups to be an important omission. These subgroup analyses were pre-specified within the statistical analysis plan and none of the reasons stated by the company appear sufficient to preclude these analyses being presented alongside those based on the overall ITT population. Indeed, it is possible that variability (i.e. differences that appear to occur between patients by chance) in the GiACTA trial results may actually be explained by observable differences in patient characteristics. The newly diagnosed and relapsed/refractory populations represent potentially important indicators of heterogeneity (i.e. difference that occur between patients that can be explained) which warrants further investigation.

Although the company reported a lack of difference in efficacy between these subgroups, the clinical and statistical basis for this conclusion is unclear. The ERG also notes that a lack of a clinically meaningful difference in efficacy between the subgroups would be evident if the cost-effectiveness results for each subgroup were similar to the results overall ITT population. However, in the absence of any cost-effectiveness results reported by the company for these subgroups, it was not possible to confirm the company's statement and/or to demonstrate that any difference which does exist across the subgroups does not lead to meaningful differences in the cost-effectiveness results.

The ERG requested analyses and results for the following subgroups: (i) newly diagnosed GCA and (ii) relapsed/refractory GCA. These additional analyses were subsequently provided by the company in response to the points for clarification. Section 5.2.10 reports the additional cost-effectiveness results provided by the company for these subgroups.

5.2.4 Interventions and comparators

The cost-effectiveness analysis was based on a comparison of two of the four treatment arms from the GiACTA trial: TCZ-QW + 26-week prednisone taper and placebo-QW + 52-week prednisone taper. The TCZ-QW dosing regimen was selected in line with the CHMP positive opinion for marketing authorisation. Although prednisone is not licensed for GCA, glucocorticoids are the mainstay of treatment for patients with GCA. The company also stated that the comparator treatment and dosing schedule is consistent with the most rapid taper regimen recommended by existing BSR/BHPR

guidelines. The company did not formally include other steroid-sparing treatments as separate comparators but noted that their use was permitted within the GiACTA trial at a stable dose.

The comparator regimen included in the model was considered by the company to appropriately reflect the final NICE scope, which simply stated that the comparator should be established treatments. The submission noted that while the 52-week prednisone tapering regimen was consistent with the most rapid taper regimen, clinicians often use a longer tapering regimen in routine clinical practice (typically 18-24 months). The submission also highlighted that a longer tapering regimen could lead to a greater cumulative steroid burden in clinical practice compared to that observed in the GiACTA trial. However, the sub mission did not discuss other issues that might affect the generalisability of the GiACTA trials results to routine clinical practice. Importantly, several studies have shown that both the initial GC dose and the tapering schedule appear to influence the relapse rate. Higher relapse rates have been reported in the context of clinical trials with adjuvant therapies where GC tapering is more aggressive than in routine clinical practice.[13] Consequently, although the 52-week tapering regimen is consistent with the most rapid tapering regimen recommended in the BSR/BHPR guidelines, uncertainty remains concerning the generalisability of this tapering regimen and the associated relapse rate to a longer GC tapering regimen (18-24 months) more conventionally used.

Clinical advice received by the ERG indicated that patients in England and Wales would be likely to be treated with prednisolone rather than prednisone. This is supported from UK data from CPRD which reported that 99.7% of GCA patients received prednisolone.[7] The current list price of prednisolone (5mg, 28 tablets = £0.81) is lower than prednisone (5mg, 30 tablets = £26.70). The ERG therefore requested further justification for assuming the cost of oral prednisone rather than prednisolone and an additional scenario assuming the lower acquisition cost of prednisolone. In their response, the company agreed that prednisolone is recommended in current guidelines and altered their costing assumptions accordingly as part of their revised model and base-case analyses. The results presented in Section 5.2.10 are based on these revised analyses.

The intervention being assessed is TCZ-QW combined with a much shorter prednisone tapering regimen (26 weeks) than routinely used in clinical practice. There exists some uncertainty whether in routine practice clinicians will follow the more rapid steroid tapering regimen alongside tocilizumab. However, clinical advice received by the ERG supported the view that clinicians would seek to taper steroids more quickly with adjuvant use of tocilizumab.

There also exist important uncertainties regarding the appropriate duration of treatment with tocilizumab. Although the GiACTA trial assessed 52-week continued treatment with TCZ-QW, the CHMP Positive Opinion for Marketing Authorisation states that TCZ-QW can be given beyond 52

weeks depending on disease activity, physician discretion, and patient choice. The company base-case analysis assumes that TCZ-QW will be used continuously for a 24-month period. The duration of treatment was justified as being consistent with the current duration of conventional steroid treatment, where clinical practice aims to withdraw therapy as early as possible without risking a GCA relapse/flare. However, in the absence of a clear stopping rule for tocilizumab there remains significant uncertainty concerning the appropriate duration of tocilizumab treatment.

The uncertainty surrounding the optimal duration of tocilizumab treatment has important implications for the cost-effectiveness results. The cost-effectiveness of continued use of tocilizumab beyond the 52-week period reported in the GiACTA trial will be significantly influenced by the uncertainty and assumptions made concerning the ongoing efficacy of TCZ-QW over longer treatment durations.

A key assumption applied in the base-case analysis is that the efficacy of tocilizumab over longer treatment durations will follow the same trend as observed in the within-trial period. Although the company presented scenario analysis for alternative fixed durations of tocilizumab treatment (between 12 and 60 months), these scenarios only address one aspect of the uncertainty; the cost implications of alternative treatment durations. As such, these scenarios only partially represent the extent of uncertainty in the cost-effectiveness results since identical efficacy is assumed across each scenario.

This uncertainty and implications for the cost-effectiveness results are further explored by the ERG in Section 6.

5.2.5 Perspective, time horizon and discounting

The perspective of the company's analysis was the NHS and Personal Social Services (NHS & PSS). The time horizon used in the model was 30 years, assumed to be equivalent to a lifetime horizon. The use of a lifetime horizon is appropriate since several GCA-related complications and GC-related adverse events have lifetime HRQoL and cost consequences. However the ERG considers that there are significant uncertainties relating to the extrapolation assumptions employed within the economic model that have not been fully addressed in the company submission.

5.2.6 Treatment effectiveness and extrapolation

The effectiveness of TCZ-QW+26 versus prednisone alone was assessed in terms of the impact on GCA control (time in remission, number of flares and GCA related complications) and the impact of steroid sparing (cumulative prednisone dose, GC related adverse events). Effectiveness data was derived from the GiACTA trial (time in remission, number of flares), external literature (GCA related complications) and real world data (GC related adverse events).

The main health state transitions, assumptions and sources are summarised in Table 8 and are described in more detail in the following sections.

Table 8: Main health state transitions

Transition	Assumption	Source
Remission to relapse/flare	Time dependent, calculated from GiACTA trial data of the time to first flare event and extrapolated over a lifetime using separate parametric survival distributions fitted to individual treatment arms.	GiACTA trial data (secondary endpoint, ITT population)
Remission (escape) to subsequent relapse/flare	Constant, calculated from GiACTA trial data based on the time at risk and number of subsequent events following a first flare event. Extrapolated over a lifetime using poisson regression.	GiACTA trial data (post-hoc subgroup analysis)
GCA-related complications from relapse/flare (vision loss and stroke)	Derived from external literature and applied to each relapse/flare event.	Luqmani et al, 2016
GC-related AEs from all states receiving GC (fractures and diabetes)	Derived from real world evidence using CPRD study to estimate the risk of AEs based on cumulative steroid dose.	Real world CPRD data
Death from any state	Mortality risk based on general population mortality with an adjustment for stroke mortality.	National statistics

Table adapted from company submission

Transition – Remission to relapse/flare

Transitions from the initial remission states (on steroid and off steroid) are estimated via time-dependent transition probabilities. These are based on separate parametric survival models fitted independently to each treatment arm using patient-level data from the ITT population of the GiACTA trial.

The use of independently fitted parametric models was justified by the company based on a visual assessment of the log-cumulative hazard plots. The plots support the use of individually fitted survival models, rather than covariate based approaches using proportional hazards (PH) or accelerated-failure time (AFT) models. Alternative parametric models were then fitted to each individual treatment arm and distributions were selected based on visual inspection and formal statistical tests using the Akaike Information Criterion (AIC).

The best fitting distributions for the ITT population with the lowest AIC were the Weibull distribution for TCZ-QW+26 and the exponential distribution for the 52-week prednisone taper regimen alone. The results of the chosen parametric models were stated to have been validated based on clinical opinion and market research.

Figure 7 illustrates the Kaplan-Meier curves for each treatment arm based on the ITT population and the resulting extrapolations based on the alternative parametric functions assigned to each treatment arm.

Figure 7: Parametric extrapolation of time to first flare and Kaplan-Meier curves (ITT population)

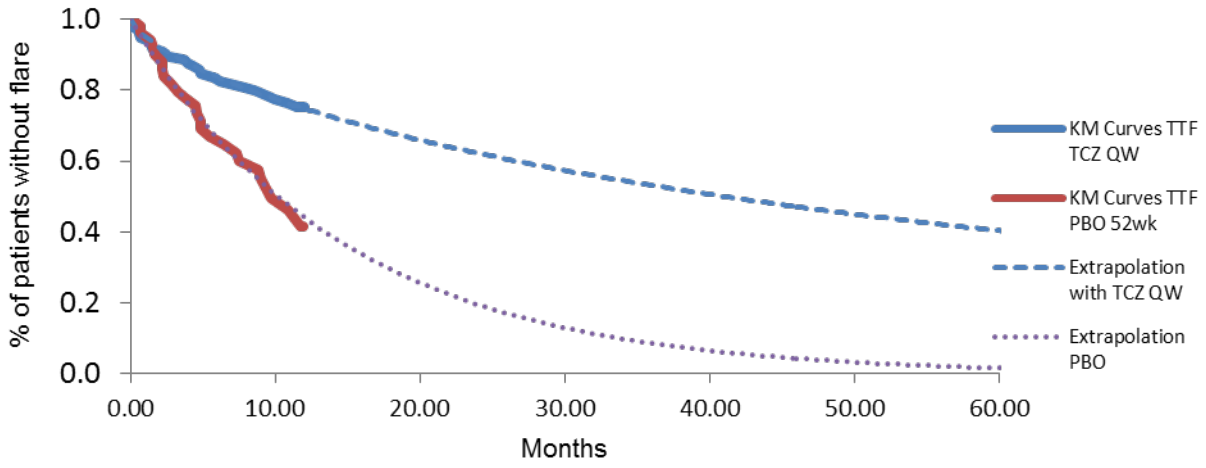


Figure replicated from CS

Figure 8 shows the longer-term predictions from the parametric function, clearly illustrating important additional gains (i.e. the area between the individual curves) are assumed beyond the discontinuation period.

Figure 8: Longer-term parametric extrapolation of time to first flare (ITT population)

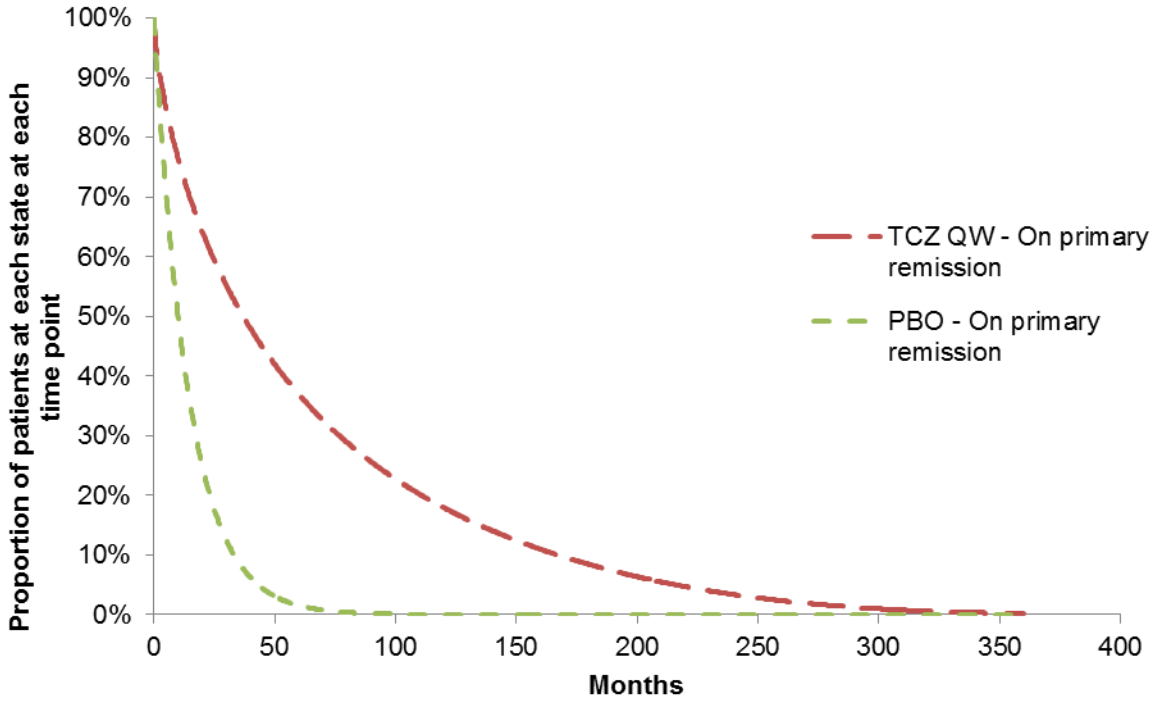


Figure replicated from CS

While fitting separate parametric models to individual treatment arms appears justifiable, it is important to note that fitting different *types* of parametric model (for example a Weibull for one treatment arm and an exponential for the other) to the separate treatment arms requires additional justification, as different models allow very different shaped distributions. Current guidance from the NICE Decision Support Unit (DSU) state that in circumstances where the proportional hazards assumption does not seem appropriate, the most sensible approach is to fit separate parametric models using the same parametric distribution allowing a two-dimensional treatment effect on both the shape and scale parameters of the parametric distribution.[32]

The ERG notes that no additional justification was provided by the company for using different types of parametric model. While the different types of distributions provides the best statistical fit to the observed data (i.e. high internal validity), the AIC tests did not indicate large differences in goodness of fit across the distributions. Furthermore, these tests do not address the external validity of the resulting extrapolations.

Table 9 summarises the goodness of fit statistics (AIC values) for each parametric distribution. The best fitting (lowest AIC) distributions for each population are highlighted by the ERG in bold: ITT population – TCZ QW+26 (Weibull), PBO+52 (Exponential); Newly diagnosed subgroup - TCZ QW+26 (Exponential), PBO+52 (Exponential) and Relapsed/refractory subgroup - TCZ QW+26 (Exponential), PBO+52 (Lognormal).

Table 9: Summary of goodness of fit statistics for time to first flare (TTFF)

	ITT population		Newly Diagnosed		Relapsed/Refractory	
	TTFF in TCZ QW + 26-wk GC taper	**TTFF in PBO QW + 52-week GC taper**	**TTFF in TCZ QW + 26-wk GC taper**	**TTFF in PBO QW + 52-week GC taper**	**TTFF in TCZ QW + 26-wk GC taper**	**TTFF in PBO QW + 52-week GC taper**
EXPONENTIAL	176.33073	**118.04365**	**85.42530**	**59.11030**	**92.89860**	60.57836
WEIBULL	**174.88006**	119.03899	85.68266	61.10613	93.19271	60.20129
LNORMAL	175.02922	118.10141	85.73792	60.33805	93.28869	**59.88400**
GAMMA	176.82294	118.10068	87.66579	62.20861	95.15233	62.08643
LLOGISTIC	174.90303	118.81808	85.71293	60.79097	93.18509	60.46400

Table replicated from company response, Table 20 p51

For each subgroup, the same parametric distributions used for the ITT population (Weibull and exponential) was applied and justified by the company based on consistency. However, while the best fitting distributions were used for the ITT population, there were alternative distributions with better statistical fits for each of the subgroups. Again, the small differences in AIC statistics do not indicate important differences in fit based on the trial period.

In circumstances where survival data require substantial extrapolation it is important to attempt to validate the predictions made by the fitted models by other means. The submission stated that the extrapolations for the ITT population were validated by comparing the proportion of patients on sustained remission to the expert clinical opinion and market research. The extrapolations were reported to be externally valid as the model output was consistent with estimates from these external sources.

The ERG identified several concerns regarding the approach and assumptions used by the company to inform the transition probabilities from the initial remission state to relapse/flare:

1) The references to expert clinical opinion and market research in the CS were unclear in relation to the associated statements of external validity. The selected parametric distribution (exponential) for the 52-week prednisone taper predicts that less than 2% of patients will not have experienced a first relapse/flare by 5 years. However, several longitudinal cohort studies of GCA patients with long term follow-up data report a significantly higher proportion of patients receiving GC that have not experienced a flare by 5 years (approximate range 30-50% across these studies).[13, 17, 29, 33] Furthermore, these studies also appear to suggest that the hazard of relapse/recurrence tends to decrease during long-term follow-up, suggesting reduced disease activity over time.[34]
2) The future trajectory of patients in the GC alone arm beyond 52-weeks is likely to follow a different trend than the period up to 52-weeks. The period up to 52-week covers the duration of the tapering period during which time patients are at highest risk of a relapse/flare event. Although patients who are successfully tapered will still face a risk of a future relapse/flare event,

inevitably these risks are likely to follow a different longer term trend than that observed during the tapering phase.

3) The assumption that patients who continue to receive TCZ beyond 52-weeks will follow a similar future trajectory as experienced during the observed follow-up period is clearly uncertain. While the Weibull distribution appears the best fitting distribution to the observed data, uncertainty exists regarding the use of this function over longer treatment durations.
4) A key assumption made in the base-case analysis is that the benefits of tocilizumab continue over a lifetime regardless of the treatment duration period. Within the economic model this is implemented by maintaining patients on the separate parametric survival function over the entire model horizon (i.e Weibull and expontential). Hence, both treatment specific and different types of parametric functions continue to be assumed over the entire extrapolation period. Consequently there is no attempt to structurally link the treatment duration period for tocilizumab to the parametric survival modelling approach. The structural disconnect means that the scenarios presented by the company concerning alternative treatment duration only consider the impact of differences in treatment costs.

This assumption that the benefits of tocilizumab treatment continue over a lifetime is justified in Table 31 of the submission on the basis that "*early results from the OLE (open label extension study) suggest that very few patients re-flare after treatment with tocilizumab*". However, Table 48 of the submission (and data reported in section B2.6.6) also state that "*50% of patients relapsed/flared after withdrawing tocilizumab therapy*". This figure appears similar to that reported by Adler et al (2016) following cessation of tocilizumab in the previous RCT, where the authors concluded that "*clinical and serologic remission in response to TCZ (tocilizumab) for 52 weeks does not result in relapse-free survival after termination of treatment*".[35]

The ERG is concerned that the assumption that the benefits of TCZ continue over a lifetime regardless of the treatment duration does not appear justifiable based on early results from the OLE study and the published results from the previous RCT. The external evidence identified by the ERG also raises uncertainties regarding the external validity of the extrapolated results for the prednisone 52-week taper.

The ERG requested further justification and evidence from the company to support the selected parametric distributions and the external validity of the longer term predictions. The company response stated that:

"*there is substantial variability between clinical opinions sought by Roche and published articles regarding the rate of flare/relapse and the time a GCA patient is at risk of these. This variability*

meant that we were unable to unanimously validate or dismiss some assumptions, nor we were able to find a suitable alternative" (Clarification response, p24) .

The ERG also requested additional justification to support the appropriateness and validity of the assumption that the benefits of tocilizumab continue over a lifetime regardless of the treatment duration and clarification. As part of the company's clarification response, they noted a number of limitations of the OLE data regarding the robustness, design and limited precision due to small numbers. The company also stated that

"*Roche recognise the duration of treatment benefit attributed to tocilizumab in the treatment of GCA patients is highly uncertain and highly impactful on the cost-effectiveness estimate. We have attempted to engage clinical opinion on this area of uncertainty, both during the dossier development and again in response to these clarification questions. However, clinical opinion varied, and clinicians were also highly uncertainty on this point*" (Clarification response, p25).

The ERG does not consider that these uncertainties have been fully addressed in the company submission or their response. These uncertainties are further explored by the ERG in Section 6.

Transition – Remission (escape) to subsequent relapse/flare

Transitions from the remission (escape) state to subsequent relapse/flare are based on constant transition probabilities. These probabilities are estimated using a Poisson regression based on a post-hoc analysis of the subgroup of patients experiencing an initial flare. The Poisson regression uses data from the time of the first flare until the end of the follow up and the observed number of subsequent flares during this period. An annualised relapse rate is estimated based on the number of flares during this period, divided by the time period (in days) and then multiplied by 365.25. These rates are then converted to weekly transition probabilities in line with the weekly model cycle.

Table 10 summarises the weekly probabilities for the ITT population and for the subgroups requested by the ERG.

Table 10: Summary of transition probabilities - Remission to relapse/flare

Population	Treatment arm	Mean rate (in log scale)	Standard Error	Mean days follow-up used within the analysis	Weekly probability of flare
ITT	Tocilizumab QW	-1.056	0.354	228	0.0106
	Placebo 52 week	-0.300	0.224	224	0.0228
Newly Diagnosed	Tocilizumab QW	-0.875	0.447	228	0.0127
	Placebo 52 week	-0.619	0.378	224	0.0166
Relapsed/ Refractory	Tocilizumab QW	-1.299	0.577	228	0.0083
	Placebo 52 week	-0.074	0.277	224	0.0285

Table replicated from company response (Table 21, p54)

In general, the results presented in Table 5 appear clinically logical in terms of the natural history. That is, the risks of subsequent flare for PBO+52 appear higher in the relapsed/refractory than the ITT and Newly Diagnosed populations. However, the ERG notes that that subgroup results report a lower absolute risk for TCZ QW+26 in the relapsed/refractory subgroup (weekly probability = 0.0083) than the equivalent risk in the newly diagnosed subgroup (0.0.127), suggesting a larger relative treatment effect in this subgroup. Although the ERG considered that a subgroup specific effect was clinically plausible, the finding that the absolute risks were lower in the tocilizumab arm of this subgroup was considered less plausible. This suggests that using subgroup specific relative effects for this transition within the model may not be appropriate. This issue is further in Section 6 by the ERG.

The CS also assumes that these transition probabilities are constant over time, suggesting that patients remain at ongoing risk of further flares for the remainder of their lifetime. A single reference was provided to support this assumption, with the company noting that flares can occur many years after initial diagnosis. The company also presented additional scenario analyses where the transition probabilities were reduced over time (5% and 10% annual reduction) recognising that many patients do not require continuous treatment.

The ERG identified further concerns regarding the approach and assumptions used by the company to inform the transition probabilities from the remission (escape) state to subsequent relapse/flare:

1. The evidence used to inform this transition is based on a post-randomisation subset of the ITT trial population. This means that the evidence used does not constitute a randomised comparison, and will be subject to confounding by both observed and unobserved covariates. This introduces

additional uncertainty and potential bias within the effectiveness estimates applied to this transition.

2. The use of a post-randomised subset also introduces an important source of selection bias. That is, the subgroup of patients who experienced a flare during the follow-up of the GiACTA trial is unlikely to representative of the entire ITT population. The prognosis of patients who relapse/flare early in the course of their treatment is likely to be different from patients who relapse/flare later. This is important because patients who did not experience a relapse/flare during the GiACTA trial follow-up period do not contribute any data to inform the transition from the remission (escape) state to subsequent relapse/flare. However, since all patients receiving prednisone alone are assumed to relapse/flare at some point during the period of extrapolation, ultimately the longer-term prognosis of all patients in the model will at some point will be informed from data entirely based on the post-randomised subset.
3. Within the CSR additional data is provided on the remission and flare status for each individual patient at each follow up assessment. The ERG reviewed these individual records and noted that there were several patients who were reported to be in 'flare at visit' during consecutive follow-up times (e.g. at weeks 44 and 48). The ERG was uncertain whether these were being treated as separate flare events or a single event within in the Poisson regression. The ERG was concerned that treating these as separate flare events might over-estimate the risk of a subsequent flare/relapse.
4. The total mean number of flares (19.67) predicted by the model over a 30-year period for the ITT population appears high for the prednisone alone comparator based on longer-term epidemiological evidence identified by the ERG. Proven et al (2003) reported a maximum of 7 flares in any *single* patient based on a median follow-up of 10-years.[17] The company model predicts of a *mean* of 10.35 relapses over the same 10-year period. Similarly, Labarca et al (2015) reported a median relapse rate of 0.4 relapses/year (IQR 0.21-0.64) over a median duration of 5-years (i.e. approximately 2 relapses over 5 years compared with the company model predictions of 5.26 over the same period).[29]

Although the ERG acknowledges that the populations included in the longer-term epidemiological evidence may be more generalisable to the newly diagnosed subgroup, the marked difference in the estimates and more general concerns regarding the impact of selection bias raise important uncertainties regarding the external validity of the model estimates.

The ERG requested further clarification on the validation undertaken and additional evidence to support the external validity of the predicted number of flares. In their response, the company noted the challenges of estimating the mean number of flares given both clinical uncertainty as well as heterogeneity in the GCA population. The company also provided additional information based on the

views of attendees (rheumatologists and ophthalmologists) from an advisory board meeting. The collective view of attendees was:

- [redacted] of GCA patients would be able to taper their GC dose over approximately [redacted]
- [redacted] of GCA patients would have a relapsing/refractory GCA which required continuous titration up and down of GCs over a period of approximately [redacted]
- [redacted] of GCA patients would require a long-term GC maintenance dose for [redacted], where their GCA was controlled at a stable dose, but attempting to withdraw GC all together would cause a flare/relapse at any time after diagnosis

(Clarification response, p25)

The collective view suggests that the disease course of the majority of patients (approx. [redacted]) can be successfully managed with conventional GC tapering durations without experiencing recurrent flare/relapse. For the remaining patients [redacted] approximately [redacted] of these will experience multiple relapses requiring a longer term GC treatment duration (3-years) and the other [redacted] require long-term GC maintenance treatment (5-years or more) due to the continued risk of flare. In contrast, the company model predicts that all GCA patients receiving conventional GC treatment will eventually experience a relapse/flare. Following this relapse, the disease is then assumed to following a chronic relapse-remitting course.

The ERG acknowledges the challenges and the heterogeneity among GCA patients. However, the collective view of the attendees appears inconsistent with the characterisation of the natural history of GCA within the company model. The ERG does not consider that these uncertainties have been fully addressed in the company submission or their response. These uncertainties are further explored by the ERG in Section 6.

The company also clarified that if a patients in a flare state for consecutive assessments (e.g. week 44 and 48) that these were counted as distinct flares. The company reported that this only affects 5 patients in the 52-week GC taper arm and no patients among the TCZ-QW arm, concluded that this was unlikely to substantially impact the cost-effectiveness calculations. However, the company did not provide an additional sensitivity analysis as requested by the ERG. The ERG's review of the CSR data identified 8 possible patients that this might affect in the 52-week GC taper arm, as opposed to 5 stated by the company. The ERG is uncertain regarding the potential impact of this assumption.

Transition – GCA-related complications from relapse/flare (vision loss and stroke)

GCA-related complications were modelled as separate events that can only be experienced by patients in the relapse/flare state. The complications included were loss of vision and stroke (fatal and non-fatal). Although these complications are rare, these were considered by the company to be the most serious and relevant GCA-related complications arising as a result of a flare/relapse.

In the absence of these complications reported in the GiACTA trial, the associated risk of these were derived from a previously published economic model comparing alternative GCA diagnostic approaches.[36] Annual incidence rates of GCA-related complications at relapse/flare (0.013% for visual loss and 0.026% for stroke) were then converted to weekly probabilities in line with the model cycle length. Approximately 40% of stroke events were assumed to be major, with a 50% mortality rate.

Table 11 summarises the probabilities of GCA-related complications assigned in the model.

Table 11: Summary of probabilities of GCA-related complications

Parameter	Value	Source
Probability of visual loss at relapse/flare	0.00025	Luqmani et al, 2016[36]
Probability of stroke at relapse/flare	0.00050	Luqmani et al, 2016[36]
Probability of minor stroke at relapse/flare	0.0030	Luqmani et al, 2016[36]
Probability of major stroke at relapse/flare	0.0020	Luqmani et al, 2016[36]
Probability of death from major stroke (in addition to background mortality from life tables)	50%	Luqmani et al, 2016[36]

As previously noted, this transition assumes a surrogate relationship between GCA-related complications and relapse/flare events and that the risks of these complications are modifiable with treatment with TCZ-QW+26. Although the use of a surrogate relationship is appropriate given the rarity of these events, the degree to which these risks are modifiable with TCQ-QW remains uncertain. An editorial by Cid and Alba (2015) reported that flares mainly occur during the first 2 years after initiation of treatment and that irreversible sight loss and ischaemic complications are unusual during controlled relapses.[34] This also appears to be reflected in the responses received by the company from their clinician advisory board, who reported the risk to be of low concern generally and easily managed for patients experiencing a flare/relapse (see Clarification response, p31).

The ERG concludes that there is uncertainty regarding the extent to which these risks can be modified by treatment with tocilizumab. However, the risk of both events included in the model is so low that their inclusion is not a significant driver of the cost-effectiveness results.

Transition - GC-related AEs from all states receiving GC (fractures and diabetes)

Given the limited number of major GC-related AE events reported in the GiACTA trial, the lifetime risks of fracture and diabetes were also derived from external evidence. These risks were estimated based on cumulative GC dose measured from the GiACTA trial and subsequently extrapolated using a logistic growth regression approach. The cumulative GC dose was then linked with the risk of fracture and diabetes based on real world evidence from CPRD.[18, 37]

The calculation of cumulative GC dose for each treatment arm was undertaken in three stages:

- Stage 1 (during initial remission): based on the alternative GC tapering regimens defined in the GiACTA trial protocol.
- Stage 2 (during secondary remission): based on separate logistic growth regressions informed by the GiACTA trial data (TCZ-QW) and real world evidence from the US Market Scan Database for the 52-week prednisone tapering regimen. The separate equations assumed that the cumulative dose over time would asymptote to a total dose of [redacted] for TCZ-QW and [redacted] for 52-week prednisone taper. The equations and associated parameter inputs are reported in Table 37 (p104) of the CS.
- Stage 3 (during relapse/flare) based on separate predictive equations of the GC dose for each treatment based on the GiACTA trial data. The equations and associated parameter inputs are reported in Table 37 (p104) of the CS.

The total cumulative GC dose calculations predicted across the 3 stages were then adjusted using CPRD real world data to ensure the predictions from the model matched the cumulative GC doses reported in the CPRD data. The company noted in their response that the CPRD data lacked complete data on daily dose and hence did not have sufficient granularity to inform the logistic growth equations used in Stage 2.

Figure 9 (replicated from Figure 12 of the CS) summarises the cumulative GC dose predicted by the cost-effectiveness model over a longer time period.

Figure 9: Cumulative GC dose predicted by the company model

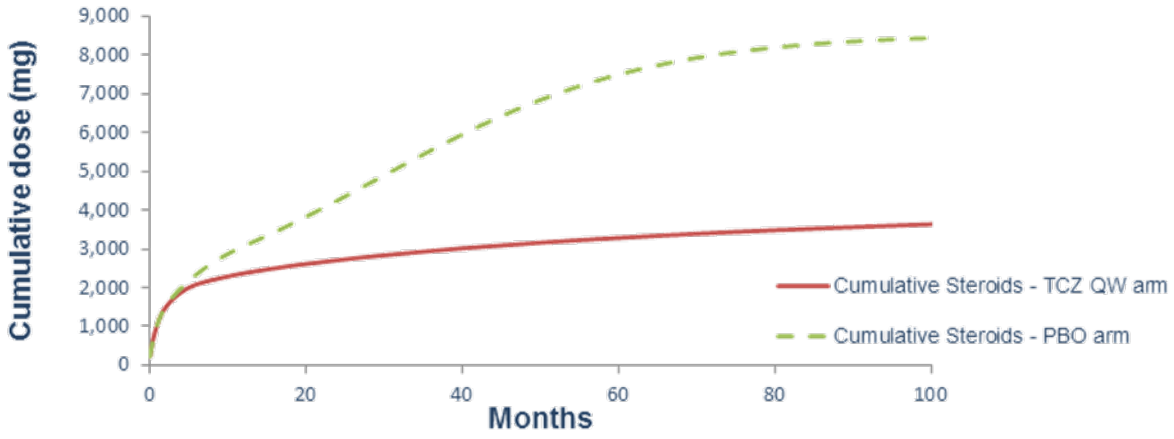


Figure replicated from CS

The ERG considers the approach to estimating cumulative dose to be reasonable and the adjustment using UK real world data increases the generalisability of the predictions. The ERG also acknowledge that the CPRD data may underestimate total GC dose as this only includes prescriptions in a primary care environment and that it was reasonable for the company to present a scenario which used the US data without further adjustment.

The ERG also notes that the same logistic growth equation and CPRD adjustment were applied across the ITT populations and subgroups. The ERG considers that the CPRD data and cumulative GC dosing is probably more reflective of the dose received for newly diagnosed patients and that higher doses, particularly in the relapsed/refractory subgroup, may be more appropriate. This issue is further explored in Section 6.

Transition - Death from any state

Estimates of background mortality applied to all states were based on 2016 UK lifetables (age and gender matched) from the Office of National Statistics, with an adjustment to avoid double counting stroke related mortality.

The ERG considers the approach to be appropriate and adequately justified by the company.

5.2.7 Health related quality of life

Remission and relapse/flare health state utilities were calculated from EQ-5D-3L (UK tariffs) data in the GiACTA trial using a mixed effects model and adjusting for baseline utility. Data were combined across all four treatment arms, given the lack of any significant differences reported between treatments and to increase the robustness of the estimates for the health state values. The company further justified this approach on the basis that the impact of a flare on a patient's quality of life was

not expected to be different across the separate arms. No time component was included as no trend in terms of utility change over time was found in the GiACTA trial data.

Table 12 summarises the main utility estimates from the mixed model for the ITT population and for the separate subgroups. The utility values estimated from the mixed model for the ITT population were 0.77 for remission and 0.64 for a relapse/flare event. The model assumes the same remission value for patients during the initial and subsequent remission periods. The lower utility estimated for a patient experiencing a relapse/flare was applied in the model for 28 days. The duration of the relapse/flare event was stated to be consistent with clinical opinion and additional analyses reported from the GiACTA trial exploring changes in utility before and after a relapse/flare.

Table 12: Summary of utilities applied to the remission and relapse states for each population

Parameter	Values for each population			Source
	ITT	Newly Diagnosed	Relapsed/Refractory	
Utility on remission	0.7713	0.8115	0.7333	GiACTA trial
Utility on flare	0.6420	0.6451	0.6343	GiACTA trial
GCA flare disutility	0.1293	0.1664	0.099	GiACTA trial

The ERG considered that the approach met the NICE reference case and that the mixed model was appropriate for the purposes of informing the model. The ERG notes that no adjustment has been made for the impact of ageing in the model and that the values for remission and flare are assumed to be constant over the entire model time horizon. However, in the absence of any significant mortality effect (i.e. other than the difference due to stroke), the ERG does not consider that this constitutes an important bias when comparing between treatment strategies in the ICER calculations.

Additional disutilities for GCA-related complications and GC-related AEs were also included and derived from the external literature. These are summarised in Table 13.

Table 13: Summary of disutilities for complications and AEs

Parameter	**Value**	**Source**
GC-related disutility	-0.07	Niederkohr and Levin 2005
GCA-related vision loss disutility from baseline	-0.36734	Luqmani et al. 2016
GCA-related minor stroke disutility from baseline	-0.17882	Luqmani et al. 2016
GCA-related major stroke disutility from baseline	-0.49122	Luqmani et al. 2016

A single GC-related disutility estimate (-0.07) is applied in the model based on an estimate reported by Niderkohr and Levin (2005).[38] This study reported the annual incidence and disutility of GC-

related adverse events based on a systematic review of previously published studies. The single GC-related distutility estimate comprises a separate disutility estimate (-0.03) applied to all patients to represent a range of common side-effects of GCs (including weight gain, 'moon-shaped' facial appearance and frequent follow-up appointments) and disutilities for less common events including fracture, psychiatric disturbance and infections which are weighted according to their incidence.

The specific disutilities and incidence of these less common events were not reported in the company submission. The ERG sourced the original values and incidence rates and a summary is presented in in Table 14. The valuation approach for each of these disutilities was not stated.

Table 14: Summary of inputs for GC-related disutility estimate

Side Effect	Disutility	Incidence (%)	Expected disutility
Base disutility	-0.03	100	-0.030
Hyperglycaemia/diabetes	-0.12	4.8	-0.006
Vertebral fracture	-0.1	6.5	-0.007
Hip/femoral fracture	-0.2	3.6	-0.007
Avascular necrosis of femoral head	-0.06	1.1	-0.001
Infection (requiring hospitalisation)	-0.19	6.7	-0.013
Peptic ulcer disease	-0.11	3.1	-0.003
Hypertension (requiring treatment)	-0.015	5.6	-0.001
Steroid myopathy	-0.05	3.4	-0.002
Psychiatric disturbance	-0.05	7.6	-0.004
Overall disutility			**-0.07**

The GC-related disutility estimate is applied for the length of the tapering period (either 26 weeks or 52 weeks) for patients in the initial remission state. Beyond the respective taper periods, no further GC-related disutility is assumed until patients experience a relapse/flare event and enter the subsequent remission state (On remission and on maintenance [escape] steroids). The GC-related disutility is then applied during each cycle patients are in the subsequent remission state. This approach assumes that following a relapse/flare event, patients will continue to incur the GC-related disutility for the remainder of their lifetime. The ERG considers that some of these disutilities do have potentially lifelong implications (e.g. diabetes, fracture). However, it may not be appropriate to continue to assume the base-disutility (-0.03) unless patients continue to receive lifelong treatment with GC.

Estimates of the distutility of GCA related complications (vision loss, minor and major stroke) were derived from a study by Luqmani et al. 2016. The valuation approach used to estimate these distutilities was not stated in the submission. Cross-checking with the source reference suggests that the disutility of visual loss were based on values estimated using a time trade-off approach. The valuation approach was not stated for stroke complications. The ERG identified minor discrepancies between several of the estimates reported in the company model and those reported in Luqmani et al. The reason for these discrepancies was unclear but the magnitude was sufficiently small that these differences were not considered likely to have any material impact on the ICER results.

5.2.8 Resources and costs

The CS provided a detailed description of resource use and costs. These related to: drug acquisition, monitoring, concomitant medication and costs related to the health states and GCA-related complications and GCA-related AEs.

The acquisition and monitoring costs of treating GCA patients with either TCZ-QW or prednisone alone are summarised in Table 15.

Table 15: Acquisition, administration and monitoring cost assumptions

Items	Intervention: Tocilizumab subcutaneous formulation	Comparator: Prednisone
Technology cost	£913.12 for 4 pre-filled syringes with 162 mg [redacted]	£26.70 for 30 tablets at 5 mg each (Following clarification, the company altered the cost data to use the lower cost of prednisolone: £0.81 for 30 tablets at 5 mg each)
Cost of treatment	The annual cost of tocilizumab treatment for a GCA patient on the weekly dosing regimen (QW) would be £11,870.56 based on list prices [redacted]. Concomitant GC treatment for the first year is modelled to be £687.06, with an additional £88.01 needed for treating flare.	The actual cost of GC treatment varies greatly for people with GCA, depending on relapse/flare or remission: a patient on maintenance treatment may have a dose as low as 5 mg/day, with the BSR Guidelines recommending up to 60 mg prednisone daily for acute relapse/flare treatment. The first year GC costs modelled for GCA patients were £885.62, with an additional £235.79 needed for treating flare.
Administration cost	Self-injection: no administration costs	Oral: no administration costs
Monitoring cost	£3 per blood test, one blood test performed every 6 weeks while on tocilizumab	Monitoring costs are associated with high-dose daily GC treatment while in relapse/flare
Tests	Not relevant	Not relevant

Replicated from company submission

The submission presented separate analyses based on the list price for tocilizumab (£913.12 for 4 pre-filled syringes with 162 mg; annual cost based on QW dosing = £11,871) and the DH/PASLU approved patient access scheme (PAS cost = [redacted] for 4 pre-filled syringes; annual cost equivalent = [redacted]).

The company acknowledged that prednisolone is recommended in current guidelines and altered their costing assumptions within their revised model and base-case. The costs of GC treatment were based on the cumulative GC dose estimated for each treatment arm.

The company submission assumes no administration costs for either tocilizumab or conventional GC treatment. However, the GiACTA clinical study report (CSR) states that the first 4 subcutaneous injections of tocilizumab required administration in a setting where medications and resuscitation facilities were available and patients were required to stay for 2 hours following each injection. The CSR also states that patients and caregivers were trained to perform the subcutaneous injection at their first visit and that clinical staff could administer the injections if a patient was unable or unwilling to self-administer.

The ERG sought further clarification from the company on possible resource use and cost implications for the NHS. The company response stated that they provide a homecare delivery and Health Check service for rheumatoid arthritis (RA) patients and hospital trusts for tocilizumab, which they are looking to continue for GCA patients. The current homecare delivery service includes up to two home visits by a qualified nurse to train the GCA patient to self-administer subcutaneous tocilizumab. The company reported that there is currently a 90% uptake of homecare delivery for RA and that the remaining 10% of patients include patients collecting them personally from the hospital pharmacy and those requiring hospital-based administration. The Health Check service is provided via the telephone and comprises up to 6 calls which includes advice and counselling where required on self-administration.

The ERG was satisfied with the company responses and assuming that these services are continued for GCA patients, the administration of TCZ seems unlikely to generate significant resource use and cost implications that were not included in company model.

Monitoring for tocilizumab requires ALT and AST levels, neutrophils and platelets and lipids to be tested every 4-8 weeks. These were assumed to be included within one blood test. A cost of £3 was derived from NHS reference costs (DAPS05 directly accessed pathology service: Haematology) and applied to all patients on tocilizumab treatment every 6 weeks.

Disease management costs were estimated separately for the following health states:

- Patients ‘on remission + on steroid’;
- Patients ‘on remission + off steroid’;
- Patients ‘on flare / relapse’;
- Patients ‘on remission + on maintenance steroids’.

Resource utilisation estimates for these states were based on data collected in the UK market research study conducted by Roche. Only limited details of this study were presented in the submission. The separate resource utilisation estimates were based on estimates of the frequency and proportion of patients expected to receive different specialist management for each state. For the different remission states, the same proportion of patients was assumed to receive care from each specialist type. However, differences in the frequency of each specialist type were assumed for the each separate remission state and for the ITT and subgroups.

Table 16 and Table 17 report the proportions and frequencies assumed by the company.

Table 16: Proportion of patients receiving specialist care in each remission state

Management Cost after diagnosis	% of patients	Cost per visit	NHS reference cost code
Rheumatologist	66%	£137	410; Rheumatology
GP	17%	£36	10.3b PSSRU 2016
Geriatrician	10%	£188	430; Geriatric Medicine
Opthalmologist	5%	£58	460; Medical Ophthalmology
Neurologist	2%	£161	400; Neurology
Other	1%	£164	300; General Medicine

Replicated from company submission

Table 17: Frequency of visits to specialist care in each remission state

Management frequency	Proportion of frequency of follow up (on remission + on steroid)			Proportion of frequency of follow up (on remission + off steroid)	Proportion of frequency of follow up (on remission + on maintenance)
	ITT	Newly Diagnosed	Relapsed/ Refractory	ITT/subgroups	ITT/subgroups
Weekly	4.6%	10.0%	4.0%	0.0%	1.9%
Every 2 weeks	14.5%	24.0%	18.0%	0.0%	9.4%
Monthly	25.9%	29.0%	29.0%	1.1%	24.5%
Every 2 months	12.7%	12.0%	14.0%	8.4%	13.2%
Every 3 months	21.0%	12.0%	22.0%	16.8%	25.9%
Every 6 months	13.0%	6.0%	9.0%	26.3%	16.5%

Replicated from company submission

The associated weekly management costs derived from the proportions and frequency estimates and applied to each state are summarised in Table 18. The ERG notes that the same weekly management costs of £26.35 were applied in the Excel model for the different populations (ITT, New-onset and Relapse/Refractory) in the 'On remission and on steroid' state, despite different frequencies reported in the previous table. The figures reported in brackets are the weekly costs estimated by the ERG based on the subgroup specific frequencies for the separate subgroups. The ERG was unclear whether this was an error or an intentional assumption made by the company. A separate deterministic sensitivity analysis has been added by the ERG at the end of this section using the subgroup specific weekly management costs for this health state.

Table 18: Weekly management costs for remission health states

Health state	Weekly management cost		
	ITT	Newly Diagnosed	Relapsed/Refractory
Patients 'on remission + on steroid'	£26.35	£26.35 (£38.41*)	£26.35 (£28.70*)
Patients 'on remission + off steroid'	£4.32	£4.32	£4.32
Patients 'on remission + on maintenance steroids'	£20.17	£20.17	£20.17

*ERG estimate

Separate proportions and frequencies were estimated for the relapse/flare state. Table 19 summarises the proportions of patients receiving care from each specialist type. The average number of appointments during the course of a flare episode was assumed to be 2.71. The weighted average cost of visits was calculated based upon the physicians involved in initial presentation and later treatment

as £259.77 in total per flare (cost of presentation = £76.11 and cost of each follow up visit = £107.40). The company also assumed that 33% of patient would receive methotrexate during the relapse/flare event.

Table 19: Proportion of patients receiving specialist care during a flare/relapse event

Management during flare	% of patients initially presenting to this speciality	% of respondents stating each physician time was involved in flare follow-up	Cost per visit	NHS reference cost code
GP	59%	44%	£36	10.3b PSSRU 2016
Rheumatologist	25%	67%	£137	410; Rheumatology [394](Department of Health 2016)(Department of Health 2016)(Department of Health 2016)
Opthalmologist	7%	10%	£58	460; Medical Ophthalmology
Geriatrician	2%	13%	£188	430; Geriatric Medicine
Neurologist	1%	6%	£161	400; Neurology
Other	7%	5%	£164	300; General Medicine

Replicated from company submission

The company submission (p121) states that "*for each resource unit cost in the economic analysis, a cost multiplier was applied to reflect that GCA patients represent high cost patients. The multiplier was calculated as 1.58 using data provided in the PSSRU 2016 by dividing the average primary care cost of the top 25% high cost patients (£381.00) over the average primary care cost of all patients (£241.00*)". The ERG notes that this multiplier does not appear to have been included within the Excel model. The reason for this discrepancy is not stated but there are several references in the submission (e.g. see response to ERG points for clarification 16) which appear to relate to assumptions and inputs included in an early model development stage and which appear to have been subsequently omitted from the final model.

The unit costs of GCA-related complications and GC-related adverse events were derived from Luqmani et al (2016) and other external sources. Table 20 summarises the unit costs. The ERG considers that these estimates appear reasonable and appropriately sourced.

Table 20: Summary of complications and adverse event costs

Event	Cost	Source
Fracture (weighted estimate based on different fracture type)	£1624 per event	Luqmani et al, 2016
Diabetes	£48.30 per week	PSSRU 2016
Vision loss- first year	£97.55 per week	Luqmani et al, 2016
Vision loss- subsequent years	£93.97 per week	Luqmani et al, 2016
Non-fatal stroke	£112.69 per week (duration =5 years)	Luqmani et al, 2016

In general, the ERG found the general presentation and reporting of the data within the submission to be difficult to follow and to validate given that the full reference to the UK market study was not provided. Further information was requested by the ERG. The company provided additional evidence and further justification which provided adequate reassurance to the ERG regarding the derivation of the numbers reported in the tables.

5.2.9 Discounting

A discount rate of 3.5% per annum was applied to both costs and outcomes in the company's base case in accordance with the NICE reference case.

5.2.10 Cost effectiveness results

As part of their clarification response, the company submitted a revised model and updated results tables. The revised submission included programming corrections requested by the ERG, alternative costing assumptions for GC (replacing the costs of prednisone with the lower acquisition costs of prednisolone) and additional subgroup analyses for newly diagnosed and relapsed/refractory GCA.

In light of the corrections and updated analyses, the ERG only reports the results presented in the revised submission and considers these to represent the relevant company base-case. In addition, since the PAS for tocilizumab already exists for other indications, the ERG only presents the PAS results and not the separate list price analysis.

The revised base-case deterministic cost-effectiveness result for the ITT population is presented in Table 21. The ICER for tociluzumab treatment with GC versus GC alone is £28,272 per additional QALY.

Table 21: Revised base-case (deterministic) cost-effectiveness results (PAS analysis)

	Total estimates				Incremental estimates				
Technologies	**Total Flares**	**Total costs (£)**	**Total LYG**	**Total QALYs**	**Incr. Flares**	**Incr. costs (£)**	**Incr. LYG**	**Incr. QALYs**	**ICER (£/QALY)**
Prednisone alone	19.12	[redacted]	12.44	8.48	12.6	£12,180	0.01	0.43	£28,272
Tocilizumab with prednisone	6.52	[redacted]	12.45	8.91					

Table 22and Table 23 present disaggregated summaries of the QALY and cost data informing the ICER estimates.

Table 22: Disaggregated summary of QALY data for base-case

	Tocilizumab	**Prednisone**	**Increment tocilizumab vs Prednisone**		
			Increment	Absolute Increment	% Absolute Increment
On Remission	8.66	7.80	0.86	0.86	200%
On Flare	0.26	0.71	-0.45	0.45	-104%
GCA-related complications	-0.01	-0.03	0.02	0.02	4%
Total QALYs	8.91	8.48	0.43	0.43	100%

The disaggregated QALY data highlights that the main driver of incremental QALY gains is the additional time patients are assumed to be in one of the remission states with tocilizumab treatment. The impact of differences due to GCA-related complications is minor. The QALY gains are conferred via two main sources: (i) a longer time to first flare which means that patients receive the higher utility of remission and avoid the utility decrement of GC-related AEs; (ii) fewer subsequent relapse/flare events meaning that a higher proportion of time, following an initial relapse/flare, is spend in the subsequent remission state.

Table 23: Disaggregated summary of cost data for base-case

	Tocilizumab	**Prednisone**	**Increment tocilizumab vs Prednisone**		
			Increment	Absolute Increment	% Absolute Increment
Tocilizumab cost	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Prednisolone cost	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Flare costs	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
GCA related costs	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
CS AE costs	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Concomitant drug	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Disease management	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Total costs	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Probabilistic sensitivity analysis

The company performed a probabilistic sensitivity analysis (PSA) where parameters were sampled probabilistically from distributions based on 1,000 simulations. The probabilistic base-case ICER reported by the company for the ITT population is £30,579 per QALY. The associated cost-effectiveness plane and acceptability curves were also presented. The probability that tocilizumab with GC is cost-effective at a threshold value of £30,000 per additional QALY is 0.59 compared with GC alone.

The ERG considers that the probabilistic ICERs represent the most appropriate estimates for the purposes of decision making. The probabilistic ICER is higher than the deterministic estimate, indicating that there are non-linearities in the model that should be accounted for in the mean ICER estimates. However, the ERG was unable to replicate the company probabilistic ICER estimates. The magnitude of variation between the company and the ERG's estimates (reported in detail in Section 6) also exceeded that which could be explained by the use of different random number sets.

The company did not separately present the mean cost and QALY estimates from the probabilistic analysis and hence the ERG could not validate or check the separate calculations informing the ICER estimates. However, the ERG believes that the company may have incorrectly calculated the probabilistic ICER by using an estimate derived from mean of the ICERs conducted within each simulation of the PSA. This approach is incorrect as the correct probabilistic ICER is the ratio based on the mean cost and QALYs derived across the simulations and not the mean ICER ratio. When calculated appropriately, by dividing the mean incremental cost across the PSA simulations by the mean incremental QALYs across the PSA, the ERG found the probabilistic ICER to be lower than the deterministic ICER (£26,748 vs £28,272 per QALY).

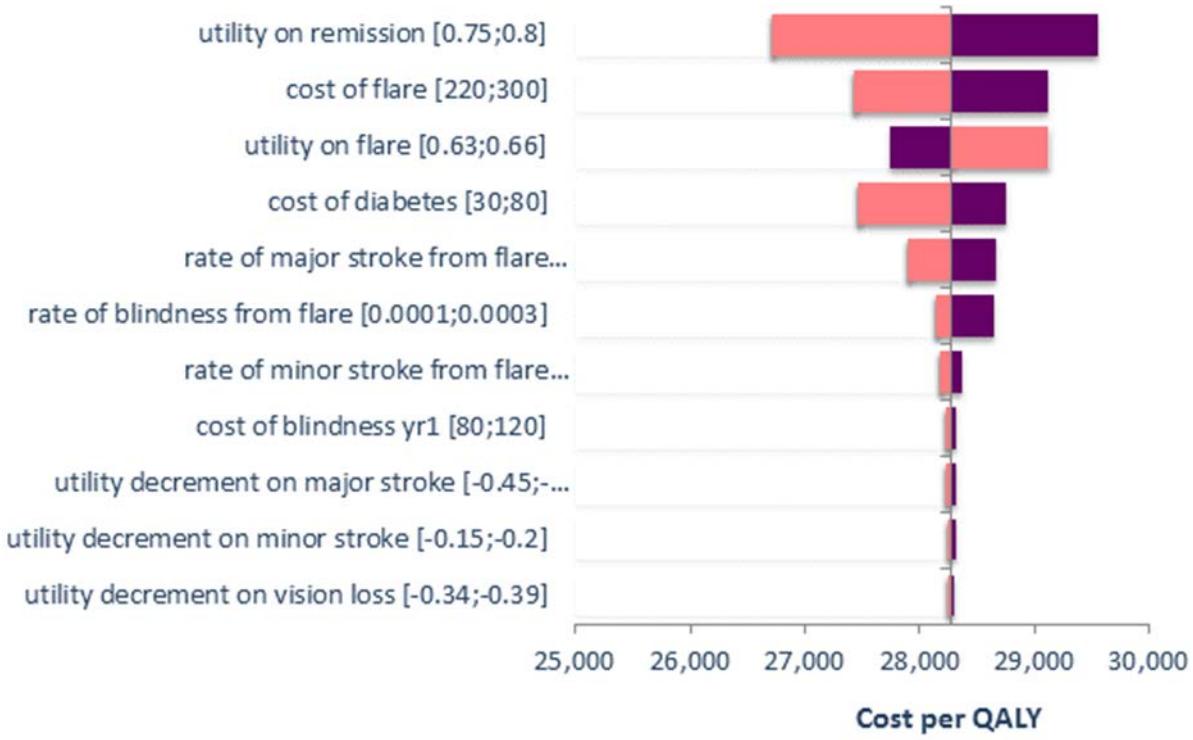
The company also provided probabilistic ICER results within their response. However, the same probabilistic ICER results reported for the ITT population (£30,579) were presented for each subgroup. Given the concerns previously noted regarding the inability to replicate the probabilistic ICER for the ITT population, the ERG presents revised probabilistic estimates for each population in Section 6.

Deterministic sensitivity analysis

The company presented a series of univariate deterministic sensitivity analyses for the ITT population to assess the impact of varying key model input parameters on the ICER. The univariate analyses were conducted by varying individual parameters across their lower and upper values based on the 10th and 90th percentile from the probabilistic distributions assigned.

Figure 10 shows a tornado diagram summarising the most influential parameters reported by the company.

Figure 10: Tornado diagram (PAS price)



CS, Figure 25 (updated sections)

The tornado diagram shows minimal variation in the ICER across the individual parameters. The highest variation was reported for the utility value assigned to the remission state with an associated ICER range between £26,711 and £29,553 per QALY. The ERG considers that it would have been

more appropriate to have used the associated 95% confidence intervals to inform the lower and upper values (i.e. 2.5th and 97.5th percentiles from the probabilistic distributions rather than the 10th and 90th percentiles) and that the results underestimate the uncertainty associated with individual parameters.

Scenarios

A range of scenario analyses were also undertaken. The alternative scenarios were presented in the CS within two separate sets of analyses. The first set of analyses assessed the use of alternative parametric models for the time to first flare and alternative stopping rules for tocilizumab (reported in Table 54, CS). The second set of analyses referred to additional scenarios considered relevant to the appraisal relating to the clinical validity and sensitivity of the inputs chosen for the base case (reported in Table 56, CS). These additional scenarios included the impact of alternative assumptions for age, the duration of tocilizumab treatment and the mean cumulative dose and variation in the rate of subsequent flares.

Table 24 summarises the results from the key scenarios across the two sets of analyses. The scenarios show that the base-case ICER appeared most sensitive to the assumptions regarding the treatment duration period and the use of the same parametric model for the time to first flare for toclllizumab as assumed for GC alone.

Table 24: Summary of key scenario analysis results – ITT population

Scenario	Scenario	Brief rationale	Impact on base-case ICER
Base case			£28,272
Age	73	Based on real world data (CPRD)	£33,159
Fixed duration of tocilizumab treatment	12 months 36 months	Uncertainty in the treatment duration period	£7,767 £47,763
Annual reduction in re-flare rate	5% 10%	Variation in the rate of re-flare reported in clinical studies	£33,902 £37,997
Mean GC cumulative dose	14g	CPRD mean dose may be underestimating actual dose due to lack of secondary care prescriptions	£25,695
Alternative parametric model (time to first flare – tocilizumab)	Exponential	Most extreme approach	£46,418

Adapted from company submission

Subgroups

Additional results were provided by the company for the newly diagnosed and relapsed/refractory subgroups as part of their response to the points for clarification. Deterministic results are provided in Table 25 and Table 26.

The ICER results were less favourable for the newly diagnosed subgroup (£37,334) and more favourable for the relapsed/refractory subgroup (£22,403), compared to the base-case ICER results for the ITT population (£28,272). The differences in the ICER estimates across the populations are driven largely by the incremental difference in the number of flares. The incremental difference in the number of flares was estimated to be -5.87 in the newly diagnosed and -19.21 in the relapsed/refractory subgroups, compared to -12.24 in the base-case ITT population. The differences across the different populations arise due to different parametric functions for the time to first flare and different rates of subsequent relapse/flare events.

Table 25: Deterministic cost-effectiveness results – Newly diagnosed subgroup

	Total estimates				Incremental estimates				
Technologies	**Total Flares**	**Total costs (£)**	**Total LYG**	**Total QALYs**	**Incr. Flares**	**Incr. costs (£)**	**Incr. LYG**	**Incr. QALYs**	**ICER (£/QALY)**
Prednisone alone	14.48	[redacted]	12.45	9.02	-5.87	£13,202	0.00	0.35	£37,334
Tocilizumab with prednisone	8.61	[redacted]	12.45	9.38					

Table 26: Deterministic cost-effectiveness results – Relapsed/Refractory subgroup

	Total estimates				Incremental estimates				
Technologies	**Total Flares**	**Total costs (£)**	**Total LYG**	**Total QALYs**	**Incr. Flares**	**Incr. costs (£)**	**Incr. LYG**	**Incr. QALYs**	**ICER (£/QALY)**
Prednisone alone	25.59	[redacted]	12.84	8.24	-19.21	£10,993	0.01	0.49	£22,403
Tocilizumab with prednisone	6.38	[redacted]	12.85	8.73					

Although separate scenario analyses were not presented in the company response for these subgroups, the ERG has repeated the same key scenarios presented for the ITT population for each subgroup. The results are summarised in Table 27 and Table 28.

Table 27: Summary of key scenario analysis results (ERG analysis) – Newly diagnosed subgroup

Scenario	Scenario	Brief rationale	Impact on base-case ICER
Base case			£37,334
Age	73	Based on real world data	£42,581
Fixed duration of tocilizumab treatment	12 months 36 months	Uncertainty in the treatment duration period	£12,354 £61,080
Annual reduction in re-flare rate	5% 10%	Variation in the rate of re-flare reported in clinical studies	£41,524 £44,450
Mean GC cumulative dose	14g	CPRD mean dose may be underestimating actual dose due to lack of secondary care prescriptions	£34,519
Alternative parametric model (time to first flare – tocilizumab)	Exponential	Most conservative approach	£71,693

Table 28: Summary of key scenario analysis results (ERG analysis) – Relapsed/refractory subgroup

Scenario	Scenario	Brief rationale	Impact on base-case ICER
Base case			£22,403
Age	73	Based on real world data	£28,093
Fixed duration of tocilizumab treatment	12 months 36 months	Uncertainty in the treatment duration period	£4,363 £39,577
Annual reduction in re-flare rate	5% 10%	Variation in the rate of re-flare reported in clinical studies	£28,708 £33,395
Mean GC cumulative dose	14g	CPRD mean dose may be underestimating actual dose due to lack of secondary care prescriptions	£20,260
Alternative parametric model (time to first flare – tocilizumab)	Exponential	Most conservative approach	£34,531

As noted in Section 5.2.8, the subgroup results reported by the company apply the same weekly management costs of £26.35 were applied in the Excel model for the different populations (ITT, Newly Diagnosed and Relapse/Refractory) in the 'On remission and on steroid' state, despite different frequencies reported. The ERG undertook a separate deterministic sensitivity analysis based on the different frequencies reported by the company. The results of these are presented in Table 29.

Table 29: ERG revised results based on alternative weekly management costs for remission health states

Weekly management cost for patients 'on remission + on steroid'		ICER	
Newly diagnosed	**Relapsed/Refractory**	**Newly diagnosed**	**Relapsed/Refractory**
£38.41*	£28.70*	£35,797	£22,253

*ERG estimate

5.2.11 Model validation and face validity check

The model was developed in-house by Roche and the face-validity of the model structure and assumptions were reported to have been reviewed by independent clinical and health economic experts. Internal validation was undertaken by an independent, external agency that performed checks on the technical programming and examined the model to identify possible logical errors or common sense issues. The external validity of the model results were also stated to have been validated by clinical opinion with explicit reference made to the re-flare rate and the proportion of patients on maintenance steroid therapy over time.

The ERG notes that while the company provided a summary of the validation steps undertaken, there was only limited detail reported in the submission on the processes and results of these validation activities. The ERG performed a series of their own independent checks of the technical programming and undertook a series of basic logical checks (e.g. altering specific inputs to determine whether the results altered in line with expectations, setting utilities to 1 to ensure that LY and QALY differences were equal etc.) to identify possible logical errors.

The ERG's logical checks identified an important error in the QALY calculations which was corrected by the company and a revised model and full set of results were provided by the company. Several other issues and concerns were also addressed by the company in their response and have been described in detail in earlier sections.

Although the ERG is satisfied with the internal validity of the model, significant uncertainties remain regarding the clinical and external validity of the longer-term extrapolations and the extent to which the current model appropriately characterises the natural history of GCA.

5.3 Conclusions of the cost effectiveness section

The ERG considered the company's economic submission to meet the requirements of the NICE reference case. However, the ERG identified a number of key uncertainties and potential errors in the CS. Several of these were subsequently addressed by the company in their response document. However, the ERG identified a number of key issues and areas of remaining uncertainty, including:

1. *Inability to replicate the probabilistic ICERs reported in the CS*

The ERG considers that the probabilistic ICERs represent the most appropriate estimates for the purposes of decision making. The ERG was unable to replicate the company's probabilistic ICER estimates and was not presented with the separate calculations used to estimate these. The estimates reported for the subgroups were also not correct.

2. *The duration of treatment and the assumption that the benefits of tocilizumab continue over a lifetime*

A key assumption applied in the base-case analysis is that the efficacy of tocilizumab over longer treatment durations will follow the same trend as observed in the within-trial period and maintained over a lifetime. Uncertainty related to the duration of treatment was explored using scenarios evaluating alternative fixed durations of tocilizumab treatment between 12 and 60 months. However, these scenarios only considered the cost implications of alternative treatment durations. The ERG considers that these scenarios do not represent the full extent of uncertainty in the cost-effectiveness results since the same efficacy is assumed within each scenario. The ERG also does not consider that the assumption that the benefits of tocilizumab continue over a lifetime regardless of the treatment duration is adequately justified or evidence based.

3. *Uncertainty concerning the choice of parametric survival models and use of different model types*

The ERG notes that no additional justification was provided by the company for using different types of parametric model based on the time to first flare. The ERG also had important concerns regarding the external validity of the longer-term predictions.

4. *Uncertainty concerning the rate of subsequent relapse/flares following an initial flare*

The CS assumes that these transition probabilities are constant over time suggesting that patients remain at ongoing risk of further flares for the remainder of their lifetime. The mean number of flares (19.67) predicted by the model over a 30-year period appears high for the prednisone alone comparator based on longer-term epidemiological evidence identified by the ERG. The use of a post-randomised subset also introduces an important source of selection bias which will impact on the validity of the longer term predictions.

Although the ERG acknowledges the challenges and the heterogeneity among GCA patients noted by the company, the ERG considers that the characterisation of the natural history of GCA and the ongoing recurrent risk of subsequent flares appears does not appear to be supported by external

evidence or the collective view of expert clinicians advising the company. The ERG does not consider that these uncertainties have been fully addressed in the scenarios presented by the company.

5. *Uncertainty regarding the generalisability of the GiACTA trial to clinical practice in England and Wales*

The ERG notes that there is an important difference in the mean age of patients in the GiACTA trial (69.05 years) and the mean age of patients in the UK CPRD data source (73 years). The ERG considered that the age reported in the UK CPRD data source more appropriately reflects the relevant population in England and Wales.

Although the 52-week tapering regimen is consistent with the most rapid tapering regimen recommended in the BSR/BHPR guidelines, there remains uncertainty surrounding the generalisability of this tapering regimen and the associated relapse rate to a longer tapering regimen (18-24 months) more conventionally used in clinical practice.

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

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

6.1 Overview

This section focuses on the additional analyses undertaken by the ERG to explore the key areas of uncertainty and concern highlighted in Section 5.

These analyses are undertaken using the revised model submitted by the company following the points for clarification. As stated in the previous section, the revised model included corrections to programming, alternative costing assumptions for GC treatment and the ability to assess the ITT populations as well as the newly diagnosed and relapsed/refractory subgroups.

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

The ERG could not replicate or validate the company's probabilistic results for their base-case analysis for the ITT population. Also, the estimates provided by the company for the separate subgroups were incorrect and reported to be the same as the ITT population. Additional simulations (1,000 iterations) were undertaken by the ERG and revised ICERs estimated by dividing the mean incremental cost by the mean incremental QALYs across the PSA.

The probabilistic results are reported in Table 30, Table 31 and Table 32 for the ITT population, newly diagnosed and relapsed/refractory subgroups.

The ERG revised probabilistic ICERs are: £26,748 (ITT population); £37,036 (new-onset) and £21,102 (relapsed-refractory). The probability that tocilizumab treatment is cost-effective at a threshold value of £30,000 per additional QALY is 0.61 (ITT population), 0.40 (new-onset subgroup) and 0.73. (relapse/refractory subgroup) compared with GC treatment alone.

Table 30: ERG revised base-case probabilistic ICER results - ITT population

	Total estimates				Incremental estimates				
Technologies	**Total Flares**	**Total costs (£)**	**Total LYG**	**Total QALYs**	**Incr. Flares**	**Incr. costs (£)**	**Incr. LYG**	**Incr. QALYs**	**ICER (£/QALY)**
Prednisone alone	20.24	[redacted]	12.42	8.44	-12.29	£12,081	0.02	0.45	£26,914
Tocilizumab with prednisone	7.95	[redacted]	12.44	8.89					

Table 31: ERG revised probabilistic ICER results - New onset subgroup

	Total estimates				Incremental estimates				
Technologies	**Total Flares**	**Total costs (£)**	**Total LYG**	**Total QALYs**	**Incr. Flares**	**Incr. costs (£)**	**Incr. LYG**	**Incr. QALYs**	**ICER (£/QALY)**
Prednisone alone	15.28	[redacted]	12.42	8.97	-5.97	£13,076	0.01	0.37	£35,766
Tocilizumab with prednisone	9.32	[redacted]	12.43	9.33					

Table 32: ERG revised probabilistic ICER results - Relapsed/refractory subgroup

	Total estimates				Incremental estimates				
Technologies	**Total Flares**	**Total costs (£)**	**Total LYG**	**Total QALYs**	**Incr. Flares**	**Incr. costs (£)**	**Incr. LYG**	**Incr. QALYs**	**ICER (£/QALY)**
Prednisone alone	26.45	[redacted]	12.82	8.19	-19.11	£10,895	0.03	0.52	£21,000
Tocilizumab with prednisone	7.34	[redacted]	12.85	8.70					

6.3 Additional ERG analyses

Although the ERG is satisfied with the internal validity of the model, significant remaining uncertainties were identified in Section 5 concerning the external validity of the longer-term extrapolations and the extent to which the current model appropriately characterises the natural history of GCA.

A series of additional scenarios were undertaken to assess the impact of these additional uncertainties and to inform an alternative ERG base-case. The alternative ERG base-case is presented in Section 6.4.

6.3.1 Scenario 1: Duration of tocilizumab treatment and benefits

The assumption that the benefits of tocilizumab continue over a lifetime regardless of the treatment duration period does not appear to be adequately justified or evidence based. An important limitation of the company base-case is the absence of any structural link between the treatment duration period and the effectiveness inputs. Consequently, the same effectiveness assumptions are employed across the separate treatment duration periods and only differences in treatment costs are assumed within the scenario analyses presented.

The longer term benefits of tocilizumab are driven by two main assumptions: (i) the continued use of treatment specific and different types of parametric functions over the entire extrapolation period for the time to first flare; and (ii) the continued use of treatment specific weekly rates of further relapse/flare events in the subsequent remission state.

The ERG notes that the model submitted by the company incorporates additional flexibility to make alternative assumptions concerning ongoing benefits beyond the treatment duration period. Additional functionality is provided in the model which allows the user to set the length of treatment duration benefit post-discontinuation of tocilizumab treatment. Within the Excel model this is reported as a 'treatment waning' parameter, allowing the user to set the length of the treatment duration benefit (in months) period post-discontinuation. The company base-case and separate scenarios set this number to a sufficiently high number (999) so that no waning of effect is assumed.

The 'treatment waning' parameter provides a potential structural link between treatment duration and benefits. At the time point at which waning is applied, patients who had previously received tocilizumab are assumed to revert to the equivalent risks as faced by patients previously treated with GC alone, albeit with different risks applied depending on whether patients are in the initial or subsequent remission state.

In the absence of robust evidence supporting a continuing effect of tocilizumab beyond the treatment period, the ERG considers that it is more appropriate to set the treatment duration benefit post-continuation to 0, such that that the longer term QALY benefits of tocilizumab treatment are more closely related to the differences predicted during treatment duration period itself.

It is important to appreciate that incorporating the waning assumption in this way does not mean that the health outcomes and costs of the alternative treatment strategies are identical in the period following discontinuation of tocilizumab treatment. Instead, the approach assumes that the differences between strategies in the post-discontinuation period arise from continuing prognostic differences due to the different distribution of patients in initial and subsequent remission health states. Since lower risks of relapse/flare events are assigned to the initial remission state compared with the subsequent remission state, the higher proportion of patients predicted to still be in the initial remission state over the treatment duration period with tocilizumab will lead to ongoing prognostic benefits in the post-treatment duration period.

The first scenario presented by the ERG (Scenario 1a) sets the treatment duration benefit post-continuation to 0 and hence applies the same risks estimated for GC patients to patients who previously received tocilizumab, depending on the state they reside at the end of the treatment period

(i.e. same exponential function for patients still in the initial remission and same weekly relapse rate for patients in the subsequent remission state).

Full probabilistic ICER results tables are presented in Table 33 (ITT population), Table 34 (newly diagnosed) and Table 35 (relapsed/refractory) for the 2-year fixed treatment duration period.

Table 33: ERG scenario 1a results - ITT population

	Total estimates				Incremental estimates				
Technologies	**Total Flares**	**Total costs (£)**	**Total LYG**	**Total QALYs**	**Incr. Flares**	**Incr. costs (£)**	**Incr. LYG**	**Incr. QALYs**	**ICER (£/QALY)**
Prednisone alone	20.13	[redacted]	12.42	8.44	-1.52	£15,992	0.00	0.12	£134,241
Tocilizumab with prednisone	18.61	[redacted]	12.43	8.55					

Table 34: ERG scenario 1a results - New-onset subgroup

	Total estimates				Incremental estimates				
Technologies	**Total Flares**	**Total costs (£)**	**Total LYG**	**Total QALYs**	**Incr. Flares**	**Incr. costs (£)**	**Incr. LYG**	**Incr. QALYs**	**ICER (£/QALY)**
Prednisone alone	15.27	[redacted]	12.38	8.94	-0.92	£15,977	0.00	0.10	£156,302
Tocilizumab with prednisone	14.35	[redacted]	12.38	9.05					

Table 35: ERG scenario 1a results - Relapsed/refractory subgroup

	Total estimates				Incremental estimates				
Technologies	**Total Flares**	**Total costs (£)**	**Total LYG**	**Total QALYs**	**Incr. Flares**	**Incr. costs (£)**	**Incr. LYG**	**Incr. QALYs**	**ICER (£/QALY)**
Prednisone alone	26.28	[redacted]	12.80	8.18	-2.05	£15,935	0.01	0.12	£127,529
Tocilizumab with prednisone	24.23	[redacted]	12.81	8.30					

The ERG probabilistic ICERs for scenario 1a are: £134,241 (ITT population); £156,302 (newly diagnosed subgroup) and £127,529 (relapsed-refractory subgroup) per QALY. Across the alternative treatment duration periods, the probabilistic ICERs ranged between: £112,806 - £166,270 (ITT population); £124,168 - £196,680 (newly diagnosed) and £108,558 - £153,437 (relapsed/refractory) per QALY.

Table 31 provides a summary for the alternative durations (between 1 to 5 years) considered in the company's scenario analysis.

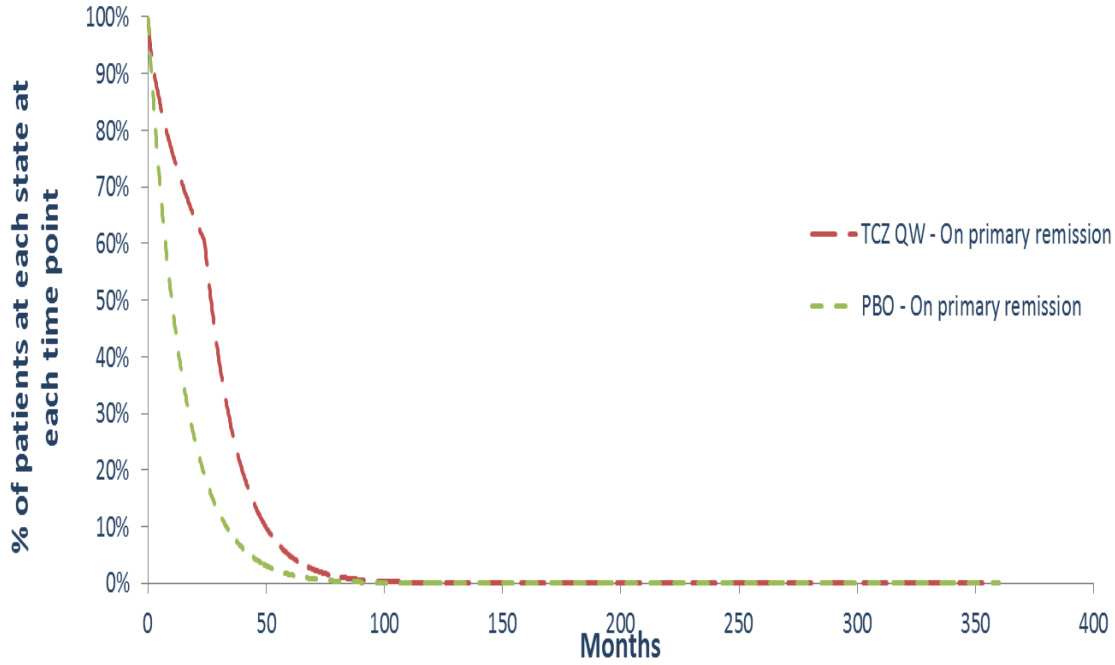
Table 36: ERG scenario 1a results - Alternative durations

	Population					
	ITT		**Newly diagnosed**		**Relapsed/refractory**	
Duration	**Incr. Flare**	**ICER**	**Incr. Flare**	**ICER**	**Incr. Flare**	**ICER**
12 months	-0.72	£112,806	-0.46	£124,168	-0.98	£108,558
24 months	-1.51	£139,122	-0.92	£156,302	-2.05	£127,529
36 months	-2.26	£147,668	-1.40	£170,429	-3.10	£138,992
48 months	-2.95	£156,573	-1.79	£181,979	-4.16	£146,923
60 months	-3.73	£166,270	-2.17	£196,680	-5.12	£153,437

The results of scenario 1a show how sensitive the ICER results are to the waning-assumption. The differences are largely driven by the much smaller incremental difference in the estimated number of flares. However, while the ERG considers that setting the treatment duration benefit post-discontinuation to 0 is more appropriate than continuing to assume treatment specific differences, the manner in which this is implemented within this scenario seems to further compound the ERG's concerns regarding the clinical plausibility and external validity of the results for GC alone. Specifically, the concerns noted regarding the high number of flares predicted now applies to both treatment strategies.

Figure 11 shows the implications for the ITT population of assuming a common parametric function for time to first flare from the point of treatment discontinuation, based on the exponential distribution used for GC alone. While the figure shows that continuing benefits are achieved post-discontinuation, the area between the curves is greatly reduced compared to the base-case analysis. More importantly, the ERG's concerns regarding the external validity of longer term predictions made for GC alone now also apply to the longer term predictions of tocilizumab. That is, a significantly higher proportion of patients are assumed to relapse and over a much shorter follow-up period compared to external epidemiological evidence.

Figure 11: Longer-term parametric extrapolation of time to first flare (ITT population): Scenario 1a



In Section 5, the ERG concluded that the future trajectory of patients in the GC alone arm beyond 52-weeks is likely to follow a different trend than the period up to 52-weeks. This is because the period up to 52-week covers the entire duration of the tapering period and represents the period over which patients are at highest risk of a relapse/flare event. The ERG questioned the relevance of this period as the basis for projecting the future probability of flare in patients who have successfully completed their taper regimen with GC alone and without experiencing a flare.

Given these concerns, the ERG considers that a more appropriate assumption would be to assume the same common parametric function for time to first flare from the point of treatment discontinuation, but to base this on the Weibull distribution from the tocilizumab arm rather than the exponential distribution from the GC alone arm. The justification for this is that the Weibull distribution is based on a decreasing risk which appears consistent with longer term epidemiological data. Furthermore, data from the tocilizumab arm may provide a better basis for subsequent projections of the future risk of GC patients who have been successfully tapered and not experienced a relapse/flare. This is because the Weibull distribution based on the tocilizumab data is informed by larger numbers of patients who: (i) didn't experience a relapse/flare and (ii) experienced longer-periods of time following the successful withdrawal of steroid treatment.

Therefore the ERG undertook a further scenario (Scenario 1b) where, at the point of tocilizumab discontinuation, patients in the GC alone treatment strategy are switched to the same Weibull function

used for tocilizumab. The results of this scenario are presented in Table 37, Table 38 and Table 39. The ERG probabilistic ICERs for scenario 1a are: £32,661(ITT population); £44,338 (newly diagnosed subgroup) and £23,730 (relapsed-refractory subgroup) per QALY.

Table 37: ERG scenario 1b results - ITT population

	Total estimates				Incremental estimates				
Technologies	**Total Flares**	**Total costs (£)**	**Total LYG**	**Total QALYs**	**Incr. Flares**	**Incr. costs (£)**	**Incr. LYG**	**Incr. QALYs**	**ICER (£/QALY)**
Prednisone alone	18.73	[redacted]	12.40	8.50	-10.95	£12,156	0.02	0.37	£32,661
Tocilizumab with prednisone	7.78	[redacted]	12.42	8.87					

Table 38: ERG scenario 1b results - New-onset subgroup

	Total estimates				Incremental estimates				
Technologies	**Total Flares**	**Total costs (£)**	**Total LYG**	**Total QALYs**	**Incr. Flares**	**Incr. costs (£)**	**Incr. LYG**	**Incr. QALYs**	**ICER (£/QALY)**
Prednisone alone	14.09	[redacted]	12.10	8.81	-5.16	£12,604	0.01	0.28	£44,338
Tocilizumab with prednisone	8.93	[redacted]	12.11	9.10					

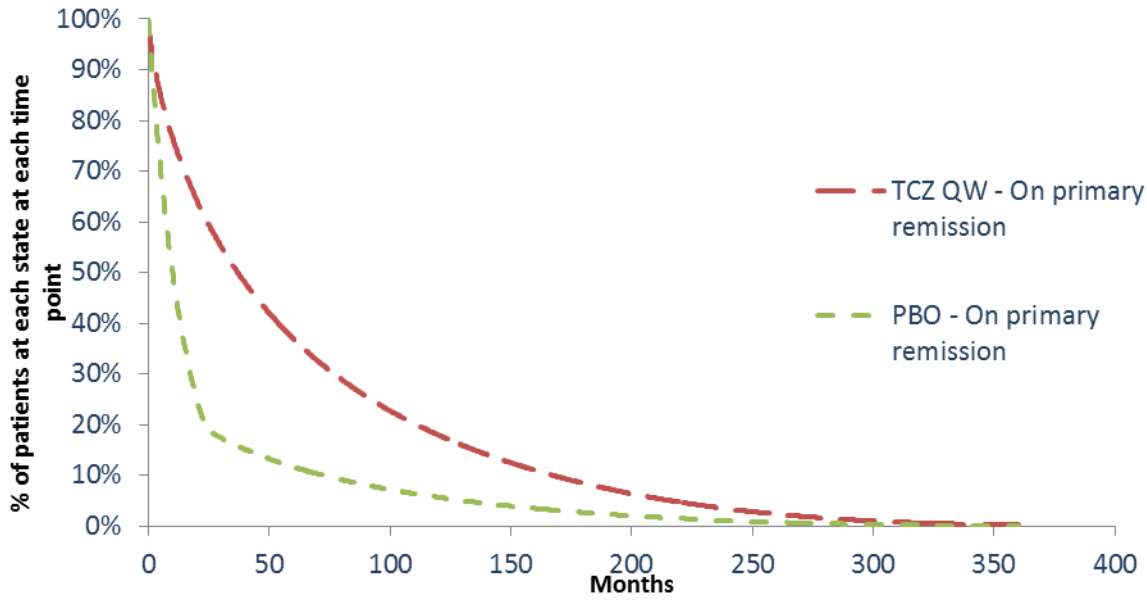
Table 39: ERG scenario 1b results - Relapsed/refractory subgroup

	Total estimates				Incremental estimates				
Technologies	**Total Flares**	**Total costs (£)**	**Total LYG**	**Total QALYs**	**Incr. Flares**	**Incr. costs (£)**	**Incr. LYG**	**Incr. QALYs**	**ICER (£/QALY)**
Prednisone alone	24.55	[redacted]	12.42	7.98	-17.42	£10,572	0.03	0.45	£23,730
Tocilizumab with prednisone	7.13	[redacted]	12.45	8.43					

Figure 12 shows the implications in the ITT population of assuming a common parametric function for time to first flare from the point of treatment discontinuation, based on the Weibull distribution used for tocilizumab treatment. Although the switch between the 2 functions creates an unrealistic kink in the survival function, the ERG considers that Scenario 1b represents a more clinically

plausible trajectory for patients receiving GC alone, with a higher proportion of patients assumed to remain in remission over a longer.

Figure 12: Longer-term parametric extrapolation of time to first flare (ITT population): Scenario 1b



The ERG notes that while this approach provides a more realistic projection for GC alone, the switch between distributions is based on an assumption of a 2-year treatment period with tocilizumab treatment. This means that the trajectory of GC alone patients is switched at 2 years i.e. 1 year after successful tapering. This means that extrapolation of the period between 12-24 months for GC alone is still being informed by data from a period over which patients are at higher risk. Consequently there remains significant uncertainty regarding the appropriate shape of the parametric distribution over longer-periods and the relevant time period over which the shapes may differ (i.e. during the initial tapering period and post tapering).

6.3.1 Scenario 2: The probability of subsequent flare

While the ERG considers that Scenario1b provides a more appropriate assumption for informing longer term projections of the time to first flare, the overall mean relapse rate remains high and appears inconsistent with longer-term epidemiological evidence identified by the ERG. In Section 5, the ERG also identified several concerns regarding the approach and assumptions used by the company to inform the transition probabilities from the remission (escape) state to subsequent relapse/flare. These concerns relate to possible selection bias and the external validity of the total number of flares predicted for the GC alone strategy.

Although several references reporting longer-term relapse data were identified by the ERG, only one study was identified which reported an annualised relapse rate over a longer time horizon. The study by Larbaca et al (2016) reported a median rate of 0.4 relapses per year (IQR 0.21-0.64) over a median duration of 5-years follow-up. The study was a retrospective cohort study of 286 patients with biopsy-proven GCA from 1998 to 2013.

The ERG considers that this study provides a useful basis to assess the external validity of the company estimates and for potentially informing the estimate of the weekly probability of flare applied to the flare state for patients receiving GC alone. A separate scenario (Scenario 2) was conducted by the ERG using this external source.

In study by Larbaca et al, patients were followed up from the point of diagnosis. Consequently, the annual relapse rate reported (0.4 per year) appear most relevant to the newly diagnosed subgroup. A series of adjustments and assumptions were made by the ERG to generalise the data across the separate populations to inform the ERG's scenario analysis:

1) A weekly probability of 0.0076 was estimated based on the annual rate of 0.4 reported. This was assumed to represent the probability of flare for the newly diagnosed subgroup for patients receiving GC alone.

2) A relative hazard between subgroups was estimated by the ERG based on the relapse rates reported for GC alone in the GiACTA trial between the newly diagnosed and the ITT and relapsed/refractory populations. This relative hazard was then applied to the annual rate of 0.4 in order to estimate equivalent rates for the ITT and relapsed/refractory populations for GC alone.

3) The relative hazards between tocilizumab treatment and GC alone were then estimated from the GiACTA trial data and applied to the population specific relapse rates estimated for GC alone. However, rather than using subgroup specific hazard ratios, the ERG used the overall ITT relative hazard for all populations, noting concerns previously highlighted in Section 5 regarding the clinical plausibility of the subgroup relative effects.

4) Weekly probabilities for each population for TCZ QW+26 and GC alone were then estimated.

Table 40 provides a comparison of the alternative ERG estimates (based on external data and GiACTA trial data) for the weekly probability of flare applied to the subsequent remission and those used in the company submission (based entirely on the GiACTA trial data).

Table 40: Comparison of weekly probability of flare applied to the subsequent remission state

Population	Treatment arm	Mean rate (in log scale)	Weekly probability of flare: ERG	Weekly probability of flare: Company
ITT	Tocilizumab QW	-1.36928	0.0049	0.0106
	Placebo 52 week	-0.59736	0.0105	0.0228
Newly diagnosed	Tocilizumab QW	-1.68821	0.0035	0.0127
	Placebo 52 week	-0.91629	0.0076	0.0166
Relapsed/ Refractory	Tocilizumab QW	-1.14328	0.0061	0.0083
	Placebo 52 week	-0.37136	0.0131	0.0285

The weekly probabilities estimated by the ERG are lower than those used in the company base-case. Importantly the ERG estimates are also logically consistent across the subgroups (i.e. the weekly probability for TCZ QW+26 is higher in the relapsed/refractory subgroup compared to the ITT and newly diagnosed populations).

Full probabilistic ICER results tables are presented in Table 41 (ITT population), Table 42 (newly diagnosed) and Table 43 (relapsed/refractory) for a 2-year fixed treatment duration period with tocilizumab. The ERG probabilistic ICERs for scenario 2 are: £39,579 (ITT population); £41,322 (newly diagnosed subgroup) and £37,582 (relapsed-refractory subgroup) per QALY.

Table 41: ERG scenario 2 results - ITT population

	Total estimates				Incremental estimates				
Technologies	Total Flares	Total costs (£)	Total LYG	Total QALYs	Incr. Flares	Incr. costs (£)	Incr. LYG	Incr. QALYs	ICER (£/QALY)
Prednisone alone	9.58	[redacted]	12.44	8.61	-5.62	£13,371	0.01	0.34	£39,579
Tocilizumab with prednisone	3.95	[redacted]	12.45	8.95					

Table 42: ERG scenario 2 results - Newly diagnosed

	Total estimates				Incremental estimates				
Technologies	**Total Flares**	**Total costs (£)**	**Total LYG**	**Total QALYs**	**Incr. Flares**	**Incr. costs (£)**	**Incr. LYG**	**Incr. QALYs**	**ICER (£/QALY)**
Prednisone alone	7.15	[redacted]	12.42	9.12	-4.07	£13,440	0.01	0.33	£41,322
Tocilizumab with prednisone	3.09	[redacted]	12.43	9.44					

Table 43: ERG scenario 2 results - Relapsed/refractory

	Total estimates				Incremental estimates				
Technologies	**Total Flares**	**Total costs (£)**	**Total LYG**	**Total QALYs**	**Incr. Flares**	**Incr. costs (£)**	**Incr. LYG**	**Incr. QALYs**	**ICER (£/QALY)**
Prednisone alone	12.28	[redacted]	12.81	8.36	-7.34	£13,084	0.01	0.35	£37,582
Tocilizumab with prednisone	4.93	[redacted]	12.82	8.71					

The ERG considers that Scenario 2 provides lower and more clinically plausible estimates of the mean number of flares over a patient's lifetime.

6.3 ERG alternative base-case

The assumptions and approaches applied by the ERG for scenarios 1b and 2 were combined and used as part of an ERG alternative base-case. Two further amendments are also proposed within the ERG alternative base-case:

1. A mean age of 73 years is assumed based on the UK CPRD data source on the basis that this more appropriately reflects the real world clinical population in England and Wales.
2. The ERG considers that the CPRD data and cumulative GC dosing is probably more reflective of the dose received for newly diagnosed patients and that higher doses, particularly for the relapsed/refractory subgroup, may be more appropriate. Therefore, the ERG excluded the CPRD adjustment applied to cumulative GC dosing for the relapsed/refractory subgroup

The results of the ERG alternative base-case for a fixed 2-year duration period for tocilizumab treatment are presented in Table 44, Table 45 and Table 46. The ERG probabilistic ICERs for are:

£65,801(ITT population); £73,046 (newly diagnosed subgroup) and £58,411 (relapsed-refractory subgroup) per QALY.

Table 44: ERG alternative base-case results (2-year treatment duration) - ITT population

	Total estimates				Incremental estimates				
Technologies	**Total Flares**	**Total costs (£)**	**Total LYG**	**Total QALYs**	**Incr. Flares**	**Incr. costs (£)**	**Incr. LYG**	**Incr. QALYs**	**ICER (£/QALY)**
Prednisone alone	7.56	[redacted]	10.76	7.51	-1.68	£14,110	0.00	0.21	£65,801
Tocilizumab with prednisone	5.88	[redacted]	10.77	7.72					

Table 45: ERG alternative base-case results (2-year treatment duration) – Newly diagnosed subgroup

	Total estimates				Incremental estimates				
Technologies	**Total Flares**	**Total costs (£)**	**Total LYG**	**Total QALYs**	**Incr. Flares**	**Incr. costs (£)**	**Incr. LYG**	**Incr. QALYs**	**ICER (£/QALY)**
Prednisone alone	5.50	[redacted]	10.51	7.78	-1.10	£13,748	0.00	0.19	£73,046
Tocilizumab with prednisone	4.40	[redacted]	10.51	7.97					

Table 46: ERG alternative base-case results (2-year treatment duration) - Relapsed/refractory subgroup

	Total estimates				Incremental estimates				
Technologies	**Total Flares**	**Total costs (£)**	**Total LYG**	**Total QALYs**	**Incr. Flares**	**Incr. costs (£)**	**Incr. LYG**	**Incr. QALYs**	**ICER (£/QALY)**
Prednisone alone	9.10	[redacted]	10.46	6.88	-2.18	£12,967	0.00	0.22	£58,411
Tocilizumab with prednisone	6.92	[redacted]	10.46	7.10					

Although the ERG considers that their revised base-case addresses several key areas of uncertainties, the remains significant uncertainty regarding the appropriate duration of treatment with tocilizumab. Both the company and ERG alternative base-case assume a fixed 2-year treatment period for tocilizumab. However, the scenarios presented by both the ERG demonstrate that the cost-effectiveness of continued use of tocilizumab beyond the 52-week period reported in the GiACTA

trial are significantly influenced by the uncertainty and assumptions made concerning the ongoing efficacy of TCZ-QW over longer treatment durations.

Inevitably, until longer-term evidence is available these uncertainties will remain. However, the ERG considered that a further set of results based on a 1-year treatment duration would provide useful additional information to inform the committee's deliberations. Specifically, this provides the most internally valid approach consistent with the treatment duration period assessed in the GiACTA trial with extrapolations based on the longer-term implications of differences in effectiveness reported over this observed follow-up period.

The ERG alternative base-case was repeated for a 1-year treatment duration period. However, the common parametric function for time to first flare (based on the Weibull distribution used for tocilizumab treatment) was applied based at 1-year treatment discontinuation.

The results of the ERG alternative base-case for a fixed 1-year duration period for tocilizumab treatment are presented in Table 47, Table 48 and Table 49. The ERG probabilistic ICERs for are: £36,960 (ITT population); £41,577 (newly diagnosed subgroup) and £30,158 (relapsed-refractory subgroup) per QALY. The more favourable ICER results compared to the 2-year treatment duration period are driven by the lower acquisition costs of tocilizumab over a shorter treatment period which reduces the incremental differences in total costs to a greater degree than the reduction in the incremental QALY differences.

Table 47: ERG alternative base-case results (1-year treatment duration) – ITT population

	Total estimates				Incremental estimates				
Technologies	**Total Flares**	**Total costs (£)**	**Total LYG**	**Total QALYs**	**Incr. Flares**	**Incr. costs (£)**	**Incr. LYG**	**Incr. QALYs**	**ICER (£/QALY)**
Prednisone alone	6.98	[redacted]	10.73	7.55	-1.03	£5,296	0.00	0.14	£36,960
Tocilizumab with prednisone	5.95	[redacted]	10.74	7.70					

Table 48: ERG alternative base-case results (1-year treatment duration) - Newly diagnosed subgroup

	Total estimates				Incremental estimates				
Technologies	**Total Flares**	**Total costs (£)**	**Total LYG**	**Total QALYs**	**Incr. Flares**	**Incr. costs (£)**	**Incr. LYG**	**Incr. QALYs**	**ICER (£/QALY)**
Prednisone alone	5.16	[redacted]	10.57	7.89	-0.65	£5,172	0.00	0.12	£41,577
Tocilizumab with prednisone	4.51	[redacted]	10.58	8.01					

Table 49: ERG alternative base-case results (1-year treatment duration) - Relapsed/refractory subgroup

	Total estimates				Incremental estimates				
Technologies	**Total Flares**	**Total costs (£)**	**Total LYG**	**Total QALYs**	**Incr. Flares**	**Incr. costs (£)**	**Incr. LYG**	**Incr. QALYs**	**ICER (£/QALY)**
Prednisone alone	8.57	[redacted]	10.54	6.99	-1.40	£4,638	0.00	0.15	£30,158
Tocilizumab with prednisone	7.16	[redacted]	10.54	7.14					

6.4 Conclusions from ERG analyses

A series of additional revisions and alternative assumptions were explored by the ERG using two main scenarios. These scenarios addressed uncertainties related to: (i) the duration of treatment and the assumption that the benefits of tocilizumab continue over a lifetime; (ii) uncertainty concerning the choice of parametric survival models for time to first flare and use of different model types and (iii) uncertainty concerning the rate of subsequent relapse/flares following an initial flare. Within these scenarios, the ERG proposed alternative assumptions and data sources which they considered had greater face validity and were more consistent with the natural history of GCA reported in longer-term epidemiological studies. These alternative approaches and data sources were then combined as part of an alternative ERG base-case analysis.

The ERG's alternative base-case presented results for alternative treatment duration periods between 1 and 2 years. The ERG ICER results were higher than those reported by the company. The ERG probabilistic ICERs for a 2-year treatment period were: £65,801(ITT population); £73,046 (newly diagnosed subgroup) and £58,411 (relapsed-refractory subgroup) per QALY. The ERG probabilistic ICERs for a 1-year treatment period were: £36,960 (ITT population); £41,577 (newly diagnosed subgroup) and £30,158 (relapsed-refractory subgroup) per QALY.

The ERG considers that the 1-year treatment period results provide the most internally valid estimates consistent with the treatment duration period assessed in the GiACTA trial. However, in the absence of a clear stopping rule for tocilizumab there remains significant uncertainty concerning the appropriate duration of tocilizumab treatment. The differences reported between the company and ERG highlight that important uncertainties remain concerning the optimal duration of tocilizumab treatment and the associated longer-term benefits.

7 End of life

Within this section, the ERG critiques relevant information regarding whether the intervention is likely to meet the end of life criteria published by NICE. It is recognised that this will be decided by the relevant NICE appraisal committee and this section may have no bearing upon their decision.

NICE end of life supplementary advice should be applied in the following circumstances and when all 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;

In the context of this assessment the end of life criteria are not applicable.

8 Overall conclusions

Evidence from a large, reasonably good quality RCT demonstrates the effectiveness of tocilizumab in achieving sustained remission, reducing the risk of flares, and reducing the GC burden. However, the treatment effect in new-onset vs relapsing patients was not fully explored, nor was the effect in patients with GCA vs LV or both. The generalisability of the trial is uncertain due to the age of patients, the ratio of cranial vs LV GCA patients, and the uncertainty regarding the taper that will be used with tocilizumab in practice

The available preliminary evidence indicates that around 30% of patients will flare once tocilizumab treatment is stopped: for sustained treatment benefit, continued treatment with tocilizumab is needed in a substantial proportion of patients.

The ERG was concerned that the assumption that the benefits of tocilizumab continue over a lifetime regardless of the treatment duration did not appear to be justifiable based on early results from the OLE study and the published results from the previous RCT. The external evidence identified by the ERG also raised uncertainties regarding the external validity of the longer-term predictions from the economic model.

The ERG alternative base-case proposes alternative assumptions and data sources which we consider have greater face validity and are more consistent with the natural history of GCA reported in longer-term epidemiological studies. The ERG alternative base-case ICER results were higher than those reported by the company.

The ERG considers that the 1-year treatment period results provide the most internally valid estimates consistent with the treatment duration period assessed in the GiACTA trial. However, in the absence of a clear stopping rule for tocilizumab there remains significant uncertainty concerning the appropriate duration of tocilizumab treatment. The differences reported between the company and ERG highlight that important uncertainties remain concerning the optimal duration of tocilizumab treatment and the associated longer-term benefits

Although the 52-week tapering regimen is consistent with the most rapid tapering regimen recommended in the BSR/BHPR guidelines, there remains uncertainty surrounding the generalisability of this tapering regimen and the associated relapse rate to a longer tapering regimen (18-24 months) more conventionally used in clinical practice.

8.1 Implications for research

Further reliable research is needed to determine the effectiveness of tocilizumab in maintaining remission in patients with GCA in the long term.

9 References

1. Weyand CM, Goronzy JJ. Giant-Cell Arteritis and Polymyalgia Rheumatica. *Ann Intern Med* 2003;139:505-15.

2. Hill CL, Black RJ, Nossent JC, Ruediger C, Nguyen L, Ninan JV, et al. Risk of mortality in patients with giant cell arteritis: A systematic review and meta-analysis. *Semin Arthritis Rheum* 2017;46:513-19.

3. Choices N. Giant cell arteritis (temporal arteritis) 2015. Available from: http://www.nhs.uk/Conditions/giant-cell-arteritis/Pages/Introduction.aspx

4. Evans JM, OTallon WM, Hunder GG. Increased Incidence of Aortic Aneurysm and Dissection in Giant Cell (Temporal) Arteritis A Population-Based Study. *Ann Intern Med* 1995;122:502-07.

5. Liddle J, Bartlam R, Mallen CD, Mackie SL, Prior JA, Helliwell T, et al. What is the impact of giant cell arteritis on patients' lives? A UK qualitative study. *BMJ Open* 2017;7:e017073.

6. Gonzalez-Gay MA, Vazquez-Rodriguez TR, Lopez-Diaz MJ, Miranda-Filloy JA, Gonzalez-Juanatey C, Martin J, et al. Epidemiology of giant cell arteritis and polymyalgia rheumatica. *Arthritis Rheum* 2009;61:1454-61. Available from: https://www.ncbi.nlm.nih.gov/pubmed/19790127

7. Petri H, Nevitt A, Sarsour K, Napalkov P, Collinson N. Incidence of giant cell arteritis and characteristics of patients: data-driven analysis of comorbidities. *Arthritis Care Res (Hoboken)* 2015;67:390-5.

8. NHS Choices. *Giant cell arteritis (temporal arteritis)*. In; 2017.

9. Borchers AT, Gershwin ME. Giant cell arteritis: a review of classification, pathophysiology, geoepidemiology and treatment. *Autoimmun Rev* 2012;11:A544-54. Available from: https://www.ncbi.nlm.nih.gov/pubmed/22285588

10. Salvarani C, Della Bella C, Cimino L, Macchioni P, Formisano D, Bajocchi G, et al. Risk factors for severe cranial ischaemic events in an Italian population-based cohort of patients with giant cell arteritis. *Rheumatology (Oxford)* 2009;48:250-3. Available from: https://www.ncbi.nlm.nih.gov/pubmed/19109317

11. Warrington KJ, Weyand CM. Giant cell arteritis and polymylagia rheumatica. In: Ball GV, Fessler BJ, Bridges SL, editors. *Oxford textbook of Vasculitis*. Oxford: Oxford University Press; 2014.

12. Butler N, Mundy J, Shah P. Aortic complications of giant cell arteritis: a diagnostic and management dilemma. *J Card Surg* 2010;25:572-81. Available from: https://www.ncbi.nlm.nih.gov/pubmed/20678106

13. Alba MA, Garcia-Martinez A, Prieto-Gonzalez S, Tavera-Bahillo I, Corbera-Bellalta M, Planas-Rigol E, et al. Relapses in patients with giant cell arteritis: prevalence, characteristics, and associated clinical findings in a longitudinally followed cohort of 106 patients. *Medicine (Baltimore)* 2014;93:194-201.

14. Walvick MD, Walvick MP. Giant cell arteritis: laboratory predictors of a positive temporal artery biopsy. *Ophthalmology* 2011;118:1201-4. Available from: https://www.ncbi.nlm.nih.gov/pubmed/21232803

15. Dasgupta B, Borg FA, Hassan N, Alexander L, Barraclough K, Bourke B, et al. BSR and BHPR guidelines for the management of giant cell arteritis. *Rheumatology (Oxford)* 2010;49:1594-7. Available from: https://www.ncbi.nlm.nih.gov/pubmed/20371504

16. Dasgupta B. Concise guidance: diagnosis and management of giant cell arteritis. *Clinical Medicine* 2010;10:381-6.

17. Proven A, Gabriel SE, Orces C, O'Fallon WM, Hunder GG. Glucocorticoid therapy in giant cell arteritis: duration and adverse outcomes. *Arthritis Rheum* 2003;49:703-8.

18. Wilson JC, Sarsour K, Collinson N, Tuckwell K, Musselman D, Klearman M, et al. Serious adverse effects associated with glucocorticoid therapy in patients with giant cell arteritis (GCA): a nested case-control analysis. *Semin Arthritis Rheum* 2017;46:819-27.

19. Alba MCMA. Sustained Remission: An Unmet Need in Patients with Giant-cell Arteritis. *The Journal of Rheumatology* 2015;42:1081-2.

20. Adizie T, Christidis D, Dharmapaliah C, Borg F, Dasgupta B. Efficacy and tolerability of leflunomide in difficult-to-treat polymyalgia rheumatica and giant cell arteritis: a case series. *Int J Clin Pract* 2012;66:906-9. Available from: https://www.ncbi.nlm.nih.gov/pubmed/22897467

21. Sciascia S, Piras D, Baldovino S, Russo A, Naretto C, Rossi D, et al. Mycophenolate mofetil as steroid-sparing treatment for elderly patients with giant cell arteritis: report of three cases. *Aging Clin Exp Res* 2012;24.

22. Quartuccio L, Maset M, De Maglio G, Pontarini E, Fabris M, Mansutti E, et al. Role of oral cyclophosphamide in the treatment of giant cell arteritis. *Rheumatology (Oxford)* 2012;51:1677-86. Available from: https://www.ncbi.nlm.nih.gov/pubmed/22627726

23. Seror R, Baron G, Hachulla E, Debandt M, Larroche C, Puechal X, et al. Adalimumab for steroid sparing in patients with giant-cell arteritis: results of a multicentre randomised controlled trial. *Ann Rheum Dis* 2014;73:2074-81. Available from: https://www.ncbi.nlm.nih.gov/pubmed/23897775

24. Hoffman GS, Cid MC, Hellmann DB, Guillevin L, Stone JH, Schousboe J, et al. A multicenter, randomized, double-blind, placebo-controlled trial of adjuvant methotrexate treatment for giant cell arteritis. *Arthritis Rheum* 2002;46:1309-18. Available from: https://www.ncbi.nlm.nih.gov/pubmed/12115238

25. Yates M, Loke YK, Watts RA, MacGregor AJ. Prednisolone combined with adjunctive immunosuppression is not superior to prednisolone alone in terms of efficacy and safety in giant cell arteritis: meta-analysis. *Clin Rheumatol* 2014;33:227-36. Available from: https://www.ncbi.nlm.nih.gov/pubmed/24026674

26. Hoffman-La Roche Ltd. F. Primary Clinical Study Report: A Phase III, multicenter, randomised, double-blind, placebo-controlled study to assess the efficacy and safety of tocilizumab in subjects with giant cell arteritis. . 2016.

27. Genentech Inc. Actemra Prescribing Information. 2017.

28. Tuckwell K, Collinson N, Dimonaco S, Klearman M, Blockmans D, Brouwer E, et al. Newly diagnosed vs. relapsing giant cell arteritis: Baseline data from the GiACTA trial. *Semin Arthritis Rheum* 2017;46:657-64. Available from: https://www.ncbi.nlm.nih.gov/pubmed/27998620

29. Labarca C, Koster MJ, Crowson CS, Makol A, Ytterberg SR, Matteson EL, et al. Predictors of relapse and treatment outcomes in biopsy-proven giant cell arteritis: a retrospective cohort study. *Rheumatology (Oxford)* 2016;55:347-56.

30. Orfanos P, Felizzi F, Harland D, Gale S, Tan D. *Assessing the comparative effectiveness of tocilizumab in giant cell arteritis within a de novo health economic model based on the GiACTA trial and the Market Scan Data*. In: ISPOR 22nd Annual International meeting. Boston, MA, US; 2017. Available from: https://www.ispor.org/ScientificPresentationsDatabase/Presentation/72186?pdfid=49647

31. Siebert U, Alagoz O, Bayoumi AM, Jahn B, Owens DK, Cohen DJ, et al. State-transition modeling: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force--3. *Value Health* 2012;15:812-20.

32. Latimer N. *Undertaking survival analysis for economic evaluations alongside clinical trials - extrapolation with patient-level data*. 2011. [cited 2017 6th October]. Available from: http://www.nicedsu.org.uk.

33. Restuccia G, Boiardi L, Cavazza A, Catanoso M, Macchioni P, Muratore F, et al. Flares in biopsy-proven giant cell arteritis in Northern Italy: characteristics and predictors in a long-term follow-up study. *Medicine (Baltimore)* 2016;95:e3524.

34. Cid MC, Alba MA. Sustained remission: an unmet need in patients with giant-cell arteritis. *J Rheumatol* 2015;42:1081-1-82.

35. Adler S, Reichenbach S, Kuchen S, Wermelinger F, Dan D, Seitz M, et al. Termination of Tocilizumab-treatment in giant cell arteritis: follow-up of patients after the RCT (ClinicalTrials.gov registration number: NCT01450137). *Arthritis Rheumatol* 2016;68.

36. Luqmani R, Lee E, Singh S, Gillett M, Schmidt WA, Bradburn M, et al. The Role of Ultrasound Compared to Biopsy of Temporal Arteries in the Diagnosis and Treatment of Giant Cell Arteritis (TABUL): a diagnostic accuracy and cost-effectiveness study. *Health Technol Assess* 2016;20:1-238.

37. Wilson JC, Sarsour K, Collinson N, Tuckwell K, Musselman D, Klearman M, et al. Incidence of outcomes potentially associated with corticosteroid therapy in patients with giant cell arteritis. *Semin Arthritis Rheum* 2017;46:650-56.

38. Niederkohr RD, Levin LA. Management of the patient with suspected temporal arteritis a decision-analytic approach. *Ophthalmology* 2005;112:744-56.

39. Department of Health. *NHS reference costs 2015 to 2016*. 2016. [cited Available from: https://www.gov.uk/government/publications/nhs-reference-costs-2015-to-2016.

10 Appendices

Appendix Table 1 Time to Remission for subjects not in Remission at Baseline, ITT Population

PBO QW + 26 Week GC Taper (n=50)		PBO QW + 52 Week GC Taper (n=51)		TCZ QW + 26 Week GC Taper (n=100)		TCZ Q2W + 26 Week GC Taper (n=49)	
Patient	Time to remission (days)	Patient	Time to remission (days)	Patient	Time to remission (days)	Patient	Time to remission (days)
A1	28	B1	15	C1	22	D1	29
A2	86	B2	28	C2	8	D2	57
A3	8	B3	197	C3	22	D3	8
A4	23	B4	22	C4	8	D4	8
A5	8	B5	15	C5	9	D5	57
A6	9	B6	29	C6	8	D6	-
A7	8	B7	8	C7	-	D7	61
A8	8	B8	29	C8	8	D8	17
A9	107	B9	15	C9	15	D9	8
A10	8	B10	85	C10	9	D10	59
A11	8	B11	9	C11	30	D11	309
A12	-	B12	8	C12	11	D12	316
A13	8	B13	141	C13	8	D13	169
A14	8	B14	8	C14	8	D14	22
A15	-	B15	57	C15	-	D15	22
A16	85	B16	58	C16	8	D16	-
A17	8	B17	-	C17	33	D17	22
A18	7	B18	113	C18	8	D18	8
		B19	169	C19	8	D19	15
		B20	8	C20	8	D20	85
		B21	56	C21	8		
		B22	15	C22	86		
		B23	22	C23	8		
		B24	84	C24	8		
		B25	15	C25	8		
				C26	8		
				C27	9		
				C28	57		
				C29	142		
				C30	8		
				C31	6		
				C32	8		
				C33	15		
				C34	8		
				C35	6		
				C36	15		
				C37	7		
				C38	8		
				C39	8		
				C40	9		
				C41	29		
				C42	8		
				C43	12		
				C44	8		

Appendix Table 2 Number of patients who completed PROs at each time point

Patients completed PRO / patients completed blinded treatment	PBO QW + 26 Week GC Taper (n=50)	PBO QW + 52 Week GC Taper (n=51)	TCZ QW + 26 Week GC Taper (n=100)	TCZ Q2W + 26 Week GC Taper (n=49)
Baseline				
SF-36 PCS	48	49	97	49
SF-36 MCS	48	49	97	49
PGA VAS	49	51	100	49
FACIT-Fatigue	50	49	99	49
EQ-5D	50	49	99	49
Week 12				
SF-36 PCS	49	51	97	49
SF-36 MCS	49	51	97	49
PGA VAS	49	51	96	49
FACIT-Fatigue	-	-	-	-
EQ-5D	49	51	96	49
Week 24				
SF-36 PCS	46	46	90	46
SF-36 MCS	46	46	90	46
PGA VAS	47	47	90	46
FACIT-Fatigue	47	49	95	46
EQ-5D	47	47	91	46
Week 36				
SF-36 PCS	44	47	85	42
SF-36 MCS	44	47	85	42
PGA VAS	46	46	87	41
FACIT-Fatigue	-	-	-	-
EQ-5D	46	46	86	41
Week 48				
SF-36 PCS	43	45	82	40
SF-36 MCS	43	45	82	42
PGA VAS	44	46	84	41
FACIT-Fatigue	45	47	81	40
EQ-5D	44	45	84	40
Week 52				
SF-36 PCS	43	45	85	39
SF-36 MCS	43	45	85	39
PGA VAS	44	43	85	40
FACIT-Fatigue	44	45	84	40
EQ-5D	44	45	85	39

Appendix Table 3 Quality Checklist for Company Model

Company submission		Reviewer's judgment (Yes/No/Unclear/NA)	Notes
Study design			
1	Was the research question stated?	Yes	
2	Was the economic importance of the research question stated?	Yes	
3	Was/were the viewpoint(s) of the analysis clearly stated and justified?	Yes	
4	Was a rationale reported for the choice of the alternative programmes or interventions compared?	Yes	
5	Were the alternatives being compared clearly described?	Yes	
6	Was the form of economic evaluation stated?	Yes	
7	Was the choice of form of economic evaluation justified in relation to the questions addressed?	Yes	
Data collection			
8	Was/were the source(s) of effectiveness estimates used stated?	Yes	
9	Were details of the design and results of the effectiveness study given (if based on a single study)?	Yes	
10	Were details of the methods of synthesis or meta-analysis of estimates given (if based on an overview of several effectiveness studies)?	NA	
11	Were the primary outcome measure(s) for the economic evaluation clearly stated?	Yes	
12	Were the methods used to value health states and other benefits stated?	Yes.	
13	Were the details of the subjects from whom valuations were obtained given?	No	
14	Were productivity changes (if included) reported separately?	NA	
15	Was the relevance of productivity changes to the study question discussed?	No	
16	Were quantities of resources reported separately from their unit cost?	Yes	
17	Were the methods for the estimation of quantities and unit costs described?	No	
18	Were currency and price data recorded?	Yes	
19	Were details of price adjustments for inflation or currency conversion given?	No	
20	Were details of any model used given?	Yes	
21	Was there a justification for the choice of model used and the key parameters on which it was based?	Yes	

Analysis and interpretation of the results			
22	Was time horizon of cost and benefits stated?	Yes	
23	Was the discount rate stated?	Yes	
24	Was the choice of rate justified?	Yes	
25	Was an explanation given if cost or benefits were not discounted?	NA	
26	Were the details of statistical test(s) and confidence intervals given forstochastic data?	No	
27	Was the approach to sensitivity analysis described?	Yes	
28	Was the choice of variables for sensitivity analysis justified?	No	
29	Were the ranges over which the parameters were varied stated?	Yes	
30	Were relevant alternatives compared? (i.e. Were appropriate comparisons made when conducting the incremental analysis?)	Yes	
31	Was an incremental analysis reported?	Yes	
32	Were major outcomes presented in a disaggregated as well as aggregated form?	Yes	
33	Was the answer to the study question given?	Yes	
34	Did conclusions follow from the data reported?	Yes	
35	Were conclusions accompanied by the appropriate caveats?	Yes	
36	Were generalizability issues addressed?	Yes	