in collaboration with:

Erasmus School of Health Policy & Management

Maastricht University

Lenalidomide with rituximab for previously treated follicular lymphoma and marginal zone lymphoma

Produced by Kleijnen Systematic Reviews Ltd. in collaboration with Erasmus University Rotterdam (EUR) and Maastricht University

Authors
Rob Riemsma, Reviews Manager, Kleijnen Systematic Reviews Ltd, UK
Thea van Asselt, Health Economist, Maastricht UMC, Netherlands
Willem Witlox, Health Economist, Maastricht UMC
Vanesa Huertas Carrera, Systematic Reviewer, KSR Ltd
Pawel Posadzki, Systematic Reviewer, KSR Ltd
Nigel Armstrong, Health Economist, KSR Ltd
Steve Ryder, Health Economist, KSR Ltd
Bram Ramaekers, Health Economist, Maastricht UMC
Sabine Grimm, Health Economist, Maastricht UMC
Xavier Pouwels, Health Economist, Maastricht UMC
Gill Worthy, Statistician, KSR Ltd
Steven Duffy, Information Specialist, KSR Ltd
Manuela Joore, Health Economist, Maastricht UMC
Jos Kleijnen, Director, KSR Ltd, Professor of Systematic Reviews in Health Care, Maastricht University

Correspondence to
Rob Riemsma, Kleijnen Systematic Reviews
Unit 6, Escrick Business Park
Riccall Road, Escrick
York, UK
YO19 6FD

Date completed 25/09/2019

Source of funding: This report was commissioned by the NIHR HTA Programme as project number STA 12/82/01.

Declared competing interests of the authors

None.

Acknowledgements

None.

Commercial in confidence (CiC) data are highlighted in blue throughout the report.

Academic in confidence (AiC) data are highlighted in yellow throughout the report.

Rider on responsibility for report

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

This report should be referenced as follows:

Riemsma R, Van Asselt ADI, Witlox W, Huertas Carrera V, Posadzki P, Armstrong N, Ryder S, Ramaekers B, Grimm S, Pouwels X, Duffy S, Joore MA, Kleijnen J. Lenalidomide with rituximab for previously treated follicular lymphoma and marginal zone lymphoma: a Single Technology Assessment. York: Kleijnen Systematic Reviews Ltd, 2019.

Contributions of authors

Rob Riemsma acted as project lead and systematic reviewer on this assessment, critiqued the clinical effectiveness methods and evidence and contributed to the writing of the report. Thea van Asselt acted as health economic project lead, critiqued the company's economic evaluation and contributed to the writing of the report. Willem Witlox, Bram Ramaekers, Sabine Grimm, Xavier Pouwels, Steve Ryder and Nigel Armstrong acted as health economists on this assessment, critiqued the company's economic evaluation and contributed to the writing of the report. Vanesa Huertas Carrera and Pawel Posadzki acted as systematic reviewers, critiqued the clinical effectiveness methods and evidence and contributed to the writing of the report. Gill Worthy acted as statistician, critiqued the analyses in the company's submission and contributed to the writing of the report. Steven Duffy critiqued the search methods in the submission and contributed to the writing of the report. Manuela Joore acted as health economist on this assessment, critiqued the company's economic evaluation, contributed to the writing of the report and provided general guidance. Jos Kleijnen critiqued the company's definition of the decision problem and their description of the underlying health problem and current service provision, contributed to the writing of the report and supervised the project.

Abbreviations

ADCC	Antibody-dependent cellular cytotoxicity
ADCP	Antibody-dependent cellular phagocytosis
AE	Adverse events
AIC	Akaike information criterion
ASCO	American Society of Clinical Oncology
ASCT	Autologous stem cell transplantation
AWMSG	All Wales Medicines Strategy Group
bd/b.i.d	Twice daily
Benda	Bendamustine
BI	Budget impact
BIC	Bayesian information criterion
BSA	Body surface area
BSC	Best supportive care
CDF	Cancer Drugs Fund
CE	Cost effectiveness
CEA	Cost effectiveness analysis
CEAC	Cost effectiveness acceptability curve
CHMP	Committee for Medicinal Products for Human Use
CHOP	Cyclophosphamide, doxorubicin, vincristine, prednisolone
CI	Confidence interval
CIC	Commercial in confidence
CR	Complete response
CrCl	Creatinine clearance
CRD	Centre for Reviews and Dissemination
CrI	Credible interval
Cru	Complete response unconfirmed
CS	Company's submission
CSR	Clinical study report
CT	Computerised tomography;
CVP	Cyclophosphamide, vincristine, prednisone
DCRR	Durable complete response rate;
DHAP	Dexamethasone, cytarabine, cisplatin
DMC	Data monitoring committee
DOCR	Duration of complete response
DOR	Duration of response
DSU	Decision Support Unit
ECOG	Eastern Cooperative Oncology Group
EFS	Event-free survival
EM	Effect modifiers
EMA	European Medicines Agency
eMIT	Electronic market information tool
EORTC	European Organisation for Research and Treatment of Cancer
EPAR	European public assessment report
EQ-5D	European quality of life-5 dimensions
ERG	Evidence Review Group
ESHAP	Etoposide, methylprednisolone, cytarabine, cisplatin
ESMO	European Society for Medical Oncology
EUR	Erasmus University Rotterdam
FACT-Lym	Functional assessment of cancer therapy: lymphoma
FBC	Full blood count
FDA	Food and Drug Administration
FL	Follicular lymphoma
FLIPI	Follicular lymphoma international prognostic index

G-CSF	Granulocyte colony-stimulating factor
GELF	Groupe d'Etude des Lymphomes Folliculaires
H	Helicobacter
HMRN	Haematological Malignancy Research Network
HR	Hazard ratio
HRQL	Health-related quality of life
HT	Histological transformation
HTA	Health technology assessment
ICER	Incremental cost effectiveness ratio
IEE	Induction efficacy population
Ig	Immunoglobin
IITT	Induction intention-to-treat population
iNHL	Indolent non-Hodgkin's lymphoma
IPD	Individual patient data
IRC	Independent review committee
ITC	Indirect treatment comparison
ITT	Intention to treat
IV	Intravenous
IVRS	Interactive voice response system;
IWGRC	International Working Group Response Criteria
KM	Kaplan–Meier
KSR	Kleijnen Systematic Reviews
LDH	Lactate dehydrogenase
LFT	Liver function tests
MAIC	Matching-adjusted indirect comparison
MALT	Mucosa-associated lymphoid tissue lymphoma
MCL	Mantle cell lymphoma
MeSH	Medical subject headings
MHRA	Medicines and Healthcare Products Regulatory Agency
MIMS	Monthly Index of Medical Specialities
mITT	Modified intention-to-treat
MRI	Magnetic resonance imaging
MZL	Marginal zone lymphoma
NA	Not applicable
NHL	Non-Hodgkin's lymphoma
NHS	National Health Services
NICE	National Institute for Health and Care Excellence
NIHR	National Institute for Health Research
NK	Natural killer
NSAID	Non-steroidal anti-inflammatory drug
NZML	Nodal marginal zone lymphoma
O	Obinutuzumab
O-Benda	Obinutuzumab in combination with bendamustine
ONS	Office for National Statistics
OR	Odds ratio
ORR	Overall response rate
OS	Overall survival
OWSA	One-way sensitivity analysis
PAS	Patient access scheme
PD	Progressive disease;
PFS	Progression-free survival
PPS	Post-progression survival
PR	Partial response
PRESS	Peer review of electronic search strategies
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses

PSA	Probabilistic sensitivity analyses
PSSRU	Personal Social Services Research Unit
PV	Prognostic variables
QALY(s)	Quality-adjusted life year(s)
QoL	Quality of life
R^2	Lenalidomide plus rituximab
R-Benda	Rituximab in combination with bendamustine
R-chemo	Rituximab plus chemotherapy
R-CHOP	Rituximab, cyclophosphamide, doxorubicin, vincristine, prednisolone
R-CVP	Rituximab, cyclophosphamide, vincristine, prednisone
R-mono	Rituximab monotherapy
RCT	Randomised controlled trial
RS	Relative survival
RTNLT	Response to next anti-lymphoma treatment
RWE	Real world evidence
SAE	Serious adverse events
SC	Subcutaneous
ScHARR	School of Health and Related Research
SCT	Stem cell transplantation
SD	Stable disease/Standard deviation
SE	Standard error
SHTAC	Southampton Health Technology Assessments Centre
SIGN	Scottish Intercollegiate Guidelines Network
SLR	Systematic literature review
SMC	Scottish Medicines Consortium
SmPC	Summary of product characteristics
SMZL	Splenic marginal zone lymphoma
STA	Single technology appraisal
TEAE	Treatment-emergent adverse event
TFR	Tumour flare reaction
TLS	Tumour lysis syndrome
ToT	Time on treatment
TSD	Technical support document
TTNLT	Time to next anti-lymphoma treatment
TTP	Time to progression
TTR	Time to response
UK	United Kingdom
WHO	World Health Organisation
WTP	Willingness to pay

Table of Contents

Table of Tables

Table of Figures

1. SUMMARY

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

The population defined in the scope is: Adults with previously treated follicular lymphoma or marginal zone lymphoma. The population in the company submission (CS) is in line with the NICE scope.

According to the company lenalidomide plus rituximab (R^2) does not currently have a UK marketing authorisation, although the Committee for Medicinal Products for Human Use (CHMP) opinion is anticipated on [redacted], and marketing authorisation is expected in [redacted]. Therefore, the relevant population for this appraisal is currently unclear. The anticipated license is as follows: Lenalidomide in combination with rituximab (anti-CD20 antibody) is indicated for the treatment of adult patients with previously treated Follicular Lymphoma (FL) or Marginal Zone Lymphoma (MZL).

The intervention (lenalidomide in combination with rituximab) is in line with the scope.

The description of the comparators in the NICE scope is as follows: rituximab monotherapy, rituximab in combination with chemotherapy, and established clinical management without lenalidomide (including but not limited to bendamustine). The NICE scope does not make a distinction in terms of patients being rituximab refractory or not. However, the CS has different comparators for rituximab refractory patients and non-rituximab refractory patients. The company's justification for this approach is because 'patients determined to be R-refractory are treated differently to the non-R-refractory population in the UK due to the availability of an alternative treatment, obinutuzumab-bendamustine, approved in the EU and recommended by NICE'

For non-rituximab refractory patients, the company included two comparators, both different types of rituximab in combination with chemotherapy:

- Rituximab plus cyclophosphamide, doxorubicin, vincristine, prednisolone (R-CHOP)
- Rituximab plus cyclophosphamide vincristine prednisolone (R-CVP)

For rituximab refractory patients, the company included one comparator:

- Established clinical management without lenalidomide, i.e. obinutuzumab in combination with bendamustine (O-Benda).

The ERG has several concerns with these comparators. Firstly, the NICE scope does not make a distinction in terms of patients being rituximab refractory or not. Therefore, the CS should have included a comparison of R^2 with rituximab monotherapy for all patients as specified in the scope. Secondly, even if the NICE committee accepts splitting the population in rituximab refractory patients and non-rituximab refractory patients, the CS should still have included a comparison with rituximab monotherapy for both populations as specified in the scope. Thirdly, the company included O-Benda as a comparator for rituximab refractory patients. However, in the response from NICE to comments on the draft scope, NICE clearly stated that "obinutuzumab in combination with bendamustine is only used as part of the Cancer Drugs Fund therefore it is not considered a relevant comparator for disease that is refractory to rituximab." (see NICE Response to comments on draft scope (page 4)). Therefore, we believe that the submission currently does not present any relevant evidence for R-refractory patients.

1.2 Summary of clinical effectiveness evidence submitted by the company

The company submission included six studies that were deemed relevant by the company. Four studies evaluated R^2, one of these was a randomised controlled trial (RCT) of R^2 versus R-monotherapy (the AUGMENT trial), the other three did not include relevant comparators according to the NICE scope.

The remaining two studies evaluated R-CHOP versus cyclophosphamide, doxorubicin, vincristine, and prednisolone (CHOP) (Van Oers et al., 2006) and O-Benda versus bendamustine monotherapy (the GADOLIN trial). The trial by Van Oers et al. (2006) was used by the company for an unanchored indirect comparison (using individual arms of different studies) of R^2 versus R-CHOP. However, the study only included rituximab-naïve patients and was therefore not representative for the UK patient population. The GADOLIN study was used by the company for an unanchored indirect comparison of R^2 with O-Benda. However, as explained in Sections 3.3 and 4.4 of this report, O-Benda is not considered by NICE to be a relevant comparator for this appraisal; therefore, this study has been ignored in this report.

In conclusion, the CS included one relevant study, for the comparison of R^2 versus R-monotherapy: the AUGMENT trial. All patients in this trial were non-R-refractory. In addition, the company performed an unanchored indirect comparison of R^2 versus R-CHOP and R-CVP, using data for R^2 from the AUGMENT trial and pooled data for R-CHOP/R-CVP from the Haematological Malignancy Research Network (HMRN) database.

The AUGMENT trial is a randomised, double-blind, multicentre, Phase III study of R^2 versus rituximab plus placebo (R-mono) in non-R-refractory patients with FL Grade 1–3a or MZL. The study was conducted across 96 sites in 17 countries. The trial did not include any patients from the UK. The primary efficacy analyses were conducted on the ITT population, defined as all randomised patients. The primary endpoint of the study was progression-free survival (PFS), as assessed by the Independent Review Committee (IRC).

Results from the AUGMENT trial show favourable results for R^2 when compared to R-mono in terms PFS with a greater median PFS ([redacted] vs. [redacted] months; hazard ratio (HR) of [redacted] (95% confidence interval (CI): [redacted]). However, there was no evidence of a difference in overall survival (OS) with a HR of 0.61 (95% CI: 0.33 to 1.13) for patients treated with R^2 compared to R-mono. At the time of the analysis the overall survival (OS) data were immature with 16 deaths on R^2 and 26 deaths on R-mono. Overall response rate (ORR) was significantly greater for R^2 compared with R-mono (78% vs. 53%; $p<0.0001$). The complete response (CR) rate was also greater for the R^2 arm compared with R-mono (34% vs. 18%; $p=0.001$). Results for R^2 versus R-mono in MZL patients were generally less favourable for R^2 than in FL patients. However, it is important to note that PFS outcomes in the MZL subgroup are difficult to interpret because of the small sample size (63 patients in total) and imbalance in baseline prognostic factors. In terms of health-related quality of life, no clinically meaningful change from baseline in the global health status/quality of life (GHS/QoL) domain of the EORTC Quality of Life Questionnaire, Core 30 (QLQ-C30) was observed across any of the post-baseline assessment visits, regardless of treatment group. Between-group differences in mean changes were small and not clinically meaningful across all assessment visits and did not differ between FL and MZL patients.

R^2 was associated with more grade 3-4 treatment-emergent adverse events (TEAEs) and serious adverse events (SAEs) when compared to R-mono, especially lenalidomide/placebo related adverse events; but rituximab-related grade 3-4 TEAEs and SAEs were also more frequent in the R^2 arm than in the R-mono arm. R^2 was also associated with more TEAEs leading to dose reductions, dose interruptions and discontinuations of lenalidomide/placebo or rituximab when compared to R-mono. Adverse events are generally the same for FL and MZL patients; however, AEs for MZL patients are based on small numbers.

The company performed three unanchored indirect comparisons, two using data from published evidence and one using data from HMRN:

- R^2 versus R-CHOP for non-rituximab refractory patients, based on comparator data from a study by Van Oers et al. 2006 comparing R-CHOP with CHOP (only the R-CHOP arm was used in the analyses).
- R^2 versus established clinical management without lenalidomide, i.e. O-Benda for rituximab refractory patients, based on comparator data from a study by Sehn et al. (2016) comparing O-Benda with bendamustine monotherapy (only the O-Benda arm was used in the analyses).
- R^2 versus pooled data for R-CHOP/R-CVP for non-rituximab refractory patients using data from HMRN.

As mentioned above, the two unanchored indirect comparisons using published evidence have been ignored in this report. R^2 versus R-CHOP, because the study by Van Oers is not representative for UK patients, and R^2 versus O-Benda because O-Benda is not a relevant comparator for this appraisal according to NICE.

Results from the remaining matching-adjusted indirect comparison (MAIC) (R^2 versus pooled data for R-CHOP/R-CVP for non-rituximab refractory patients using data from HMRN) show a significant improvement for R^2 in OS (HR = [redacted] (95% CI: [redacted]) and time to next anti-lymphoma treatment (TTNLT, HR = [redacted] (95% CI: [redacted]) compared to R-CHOP/R-CVP, but no evidence of a difference in PFS (HR = [redacted] (95% CI: [redacted]).

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

The CS and response to clarification provided sufficient details for the ERG to appraise the literature searches conducted as part of the systematic review to identify clinical effectiveness studies. A good range of databases and resources were searched. The searches did not include study design filters in order to identify both efficacy and safety evidence. Searches conducted in September 2017 were reported, but need not have been as they were subsequently replaced by searches conducted in April 2019.

The results of the MAIC should be treated with a high degree of caution due to the fact that potentially important covariates were excluded from the matching models, small sample sizes and assumptions about the equivalence of R-CHOP and R-CVP in the HMRN data and differences in PFS definitions and length of follow-up between the two data sources, The analysis also used an unanchored MAIC involving two single treatment arms from different studies, as there was no relevant comparative trial data. This analysis makes the assumption that all effect modifiers and prognostic factors are accounted for in the model, which in practice is difficult to achieve as, in this case, one or both studies do not measure a specific variable.

1.4 Summary of cost effectiveness evidence submitted by the company

The company conducted searches for cost effectiveness, health-related quality of life and healthcare resource use evidence.

The company developed a cohort-level partitioned survival model (PSM), programmed in Excel, with three health states: progression-free (PF), post-progression (PP) and death. All patients start 'on treatment' in the PF health state. Subsequently, patients either remain on treatment or come off treatment before progressing or dying per cycle. Within PP, patients can have a treatment-free interval before receiving subsequent therapy. Patients in the PP on treatment health state remain in this health state until they die. The time horizon was lifetime and cycle length 28 days. R^2 does not currently have a UK marketing authorisation, but the patient population considered in the model is in line with the proposed license: adult patients with previously treated FL or MZL. Due to the similar prognosis of FL

and MZL patients, and the difficulty in sourcing MZL-specific data, FL and MZL populations were pooled throughout the economic analysis. The R^2 dosing regimen within the model is lenalidomide 20 mg orally once daily on days 1–21 of repeated 28-day cycles for up to 12 cycles of treatment. Rituximab is given as 375 mg/m2 every week in Cycle 1 (days 1, 8, 15 and 22) and day 1 of every 28-day cycle for Cycles 2–5. This is in line with the recommended dose in the summary of product characteristics (SmPC). Based on expert opinion, the company compared R^2 in the non-R-refractory population with R-CHOP and R-CVP, and in the R-refractory population to O-Benda.

The main source of evidence on treatment effectiveness used for intervention and comparators was the AUGMENT study for R^2 and HMRN data for R-CHOP and R-CVP. The AUGMENT study contained a mixed MZL/FL population, HMRN contained only data on FL patients. The company assumed efficacy of R-CHOP and R-CVP to be similar, therefore HMRN data for R-CHOP and R-CVP were pooled. For the economic model, this implied that the comparisons of R^2 vs. R-CHOP and R-CVP had identical outcomes for effectiveness and only differed with respect to costs.

Parametric survival curves were fitted to the matched patient level data from AUGMENT and HRMN and were then used to extrapolate survival beyond study follow-up. Survival analysis was performed for OS, PFS, TTNLT, and ToT (time on treatment). PFS and ToT data were used to determine the number of patients staying in the PF (on and off treatment) health states. PFS, TTNLT and OS data were used to determine the number of patients transitioning to the PP (on and off treatment) health states. The number of patients transitioning to the death state was derived using OS data. The curves were adjusted for treatment waning, which was assumed to occur at five years. After this time point, the comparator hazard of progressing or dying was applied to the R^2 arm. Any implausible curve crossings (for instance, OS crossing PFS) were corrected for.

For the R^2 versus R-CHOP and R-CVP comparisons, the company selected a Weibull distribution to extrapolate OS, mainly based on a previous single technology appraisal (STA). For the R-mono comparison, which was added upon request of the ERG in the response to clarification, the company chose Weibull for OS as well.

The company decided to model the PFS for R^2 versus R-CHOP/R-CVP using the Kaplan–Meier (KM) data until the maximum follow-up of 46.7 months, and applied the comparator hazard to extrapolate further. In this way, the company stated in the CS, the relative treatment effect of R^2 vs. R-CHOP/R-CVP based on the MAIC was accurately reflected. For the R-CHOP/R-CVP arm, a generalised gamma was chosen. For the R-mono comparison, a simpler approach was taken, using log-logistic distributions for both arms.

Based on the Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC), the exponential distribution best fitted the R^2 data, and the log-normal distribution fitted best to R-CHOP/R-CVP. However, as the exponential distribution would result in crossing of PFS and TTNLT around seven years, the company chose the log-normal distribution for the base-case analysis for both arms. For the R-mono comparison, the generalised gamma was used for both arms.

ToT data were used to determine the proportion of patients on treatment to calculate overall drug costs. Parametric survival curves were fitted to the ToT data which, however, produced a poor fit. Therefore, the company chose to use the KM data directly in the model, and maximum treatment durations were used to cap ToT. For the R-mono comparison the same approach was used.

The main sources of evidence on treatment-related adverse events used for intervention and comparators were the AUGMENT and RELEVANCE trials, because of a lack of safety data from

HMRN. RELEVANCE is a Phase III study comparing R^2 with R-chemotherapy for patients with previously untreated FL. AUGMENT was used for R^2, and RELEVANCE was used for R-CHOP and R-CVP, after adjusting for any possible differences in R^2 AEs between AUGMENT and RELEVANCE. In a scenario, AEs for R-CHOP/R-CVP were taken from van Oers et al. (2006) which concerned a relapsed/refractory population. Furthermore, AE incidence for maintenance treatment and autologous stem cell transplant (ASCT) were also considered.

Utility values were estimated for the health states PF, and PP off and on treatment using European quality of life-5 dimensions-3 level (EQ-5D-3L) data collected in AUGMENT. A covariate selection process was used to select the appropriate mixed effects utility model as input for the economic model. The utility values resulting from the mixed effects model were used to inform the health states in the model for all treatments, and utility values from the literature were tested in scenario analyses. Disease characteristics that were used to derive utility values from the mixed effects model were population-dependent, and therefore, the utility values for R^2 versus R-CHOP/R-CVP and R^2 versus R-mono were slightly different. The mean utility values for post-progression based on the AUGMENT trial data were higher than values from the studies identified in the systematic literature review (SLR). Utility decrements for grade 3 and 4 AEs were applied in the model for the expected duration of each AE, based on literature and previous appraisals.

The cost categories included in the model were costs associated with treatment (drug acquisition costs including subsequent therapies, drug administration costs including subsequent therapies, costs associated with treatment-related AEs), disease monitoring costs and costs associated with end of life care. For lenalidomide, dosing data had been taken directly from AUGMENT (non-R-refractory population) to align the drug costs with the efficacy data because according to the company, dose reductions for lenalidomide can occur. In the economic model, the company applied ASCT to [redacted]% of patients in R-CHOP. For R-CVP and R^2, 0% ASCT was applied as it was considered unlikely that these treatments would be used as an induction regimen prior to ASCT. Subsequent treatments were included in the model as an average one-off cost to patients entering the PP (on treatment) health state, derived using TTNLT data. Costs for patients in the R^2 arm were derived from subsequent treatments from AUGMENT. The total subsequent treatment data from the pooled R-chemotherapies in the HMRN database were used for R-CHOP and R-CVP.

Total life years (LYs) and quality adjusted life years (QALYs) gained and total costs were larger for R^2 than for R-CHOP, R-CVP an R-mono. The incremental cost effectiveness ratio (ICERs) amounted to respectively £11,471, £16,814 and £22,580 per QALY gained. Compared with the deterministic results, the PSA with 1,000 iterations showed lower incremental QALYs and costs, which resulted in increased ICERs of £13,443 and £20,896 and £26,116 respectively. The cost effectiveness acceptability curves in the economic model showed that R^2 respectively had a 82%, 72% and 69% probability of being cost effective at a willingness-to-pay (WTP) threshold of £30,000. Deterministic sensitivity analyses (DSAs) were performed by varying key model parameters to the upper and lower limits of their respective confidence intervals, but in none of these analyses the ICER exceeded the £30,000 threshold.

The company performed internal validity checks using AdVISHE and made face validity checks on model structure and other assumptions within an advisory board. External validation with data from AUGMENT showed that PFS, OS and TTNLT at one year for patients treated with R-CHOP/R-CVP were under-estimated in the model compared with the observations.

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

Separate sets of searches were conducted to identify cost effectiveness studies, health-related quality of life studies and healthcare resource use evidence. The CS provided clear, transparent and reproducible searches. A good range of databases and additional resources were searched.

The company submission was largely in line with the NICE reference case. The CS did partly deviate from the scope, however, where it concerned the comparators modelled.

The company used a PSM instead of a state transition model (STM), justified by a lack of data for relevant comparators. Although the ERG recognises the potential limitations of a STM, a PSM has several limitations related to the extrapolation, as mentioned in NICE DSU TSD 19. The ERG requested a scenario analysis using a STM as a scenario, as recommended in TSD 19, which the company did not deliver. The company clarified that while FL and MZL populations were pooled, all evidence of the comparators was based on datasets that only contained patients with FL, while the AUGMENT trial contained patients with FL and MZL. In response to questions from the ERG, the company provided additional analyses on AUGMENT trial data that showed the impact of histology on the results were not statistically significant. The company provided a FL-only scenario analysis upon request of the ERG. The ERG also requested an analysis with R-mono as a comparator, as listed in the final scope, which the company provided. O-Benda was not included in the ERG report as NICE has explicitly stated it is not considered a relevant comparator for disease that is R-refractory.

A main concern of the ERG was the questionable trustworthiness of R^2 efficacy resulting from the indirect comparison, which seemed to be inflated relative to the direct comparison data from AUGMENT. Although the ERG did not have the necessary data to quantify this uncertainty, it may have lowered the ICER substantially, favouring R^2.

The ERG had concerns about the way survival curves were selected. Although the company proposed a systematic approach of selecting the appropriate curves, there were many deviations from this systematic approach in the actual selection process. The choice of OS curve was mainly based on a previous STA (TA137: Rituximab for the treatment of relapsed or refractory stage III or IV follicular non-Hodgkin's lymphoma). In particular the choice of PFS curves was not sufficiently justified and appeared sub-optimal, with a likely overestimation of PFS in the R^2 arm, and substantial underestimation of PFS in the first year for R-CHOP and R-CVP. This matter was exacerbated by the high utility values for all health states. The ERG considered these to be potentially overestimated, being higher than or comparable to those in the general population. With utilities remaining high throughout the model, any adjustment in survival curves only had little impact on the ICER, as a high utility post-progression implied there was hardly any penalty on progression in terms of cost effectiveness. This was demonstrated in the ERG scenario analyses, where the ERG base-case in combination with a lowered utility score post-progression had the highest impact on the ICER, increasing it to £33,626 and £47,281 for R^2 versus R-CHOP and R-CVP respectively, when using the lowest value from the literature for the post-progression utilities.

The ERG questioned the applicability of AE incidences taken from a previously untreated population for the present STA, and feels it is important to seriously consider the scenario provided by the company with data from a relapsed/refractory population. Therefore, the ERG included this as one of their scenarios. Also, the ERG considered it to be inconsistent that AEs related to subsequent ASCT and R-mono therapy were only taken into account for R-CHOP and R-CVP and so this was corrected for in the ERG base-case.

The ERG also had concerns about the high utility values for the PF and PP health states, and the modest utility decrement for disease progression. Utility values for the PF and PP health states were higher than the utility reported for the general population, which seems quite unlikely in patients with treated FL or MZL. Furthermore, the ERG judges that a larger utility difference between PF and PP health states would be more plausible, and explored this in a scenario analysis using lowered utility values taken from published studies for both PP health states. For R^2 versus R-CHOP and R-CVP, this substantially increased the ICER, while for R^2 versus R-mono the ICER decreased.

The ERG questioned the company's choice to include subsequent treatments as a one-off cost to those patients entering the PP on treatment health state. The company costed for observed incidences of subsequent treatments from the data sources, which for R^2 had a much shorter follow-up than for R-CHOP/R-CVP and therefore may not be reflective of clinical practice. Furthermore, subsequent treatment costs for R-CHOP and R-CVP were, in contrast to the treatment effectiveness, calculated based on the pooled R-chemotherapies data from HMRN instead of the HMRN R-CHOP/R-CVP cohort. The ERG changed this in its base-case but the impact on the ICER was modest. In addition, the company assumed the percentage of post-induction (but pre-progression) ASCTs in R^2 to be zero, because it was not protocolised in AUGMENT and clinicians considered it unlikely that patients would receive ASCT post R^2. The ERG would have liked to see a scenario using observed frequencies, as clinical practice may sometimes contrast with protocols and clinical opinion. A non-zero observed frequency would increase the ICER for R^2 versus R-CHOP.

The ERG had some comments about the PSA, which did not enable a fully incremental analysis for more than two comparators, nor representation of multiple comparators in the cost effectiveness acceptability curve (CEAC). Furthermore, probabilistic QALYs were lower compared to the deterministic QALYs in the company base-case, likely caused by non-linearity of the model. An additional scenario analysis for the FL-only population was provided by the company in response to clarification, resulting in ICERs of £15,909 and £23,746 for the R-CHOP and R-CVP comparisons, respectively, making it the most influential scenario. For the R-mono comparison, using FL-only data lowered the ICER to £20,310.

Internal validation of the model was performed to a good standard. It is not clear whether all assumptions and extrapolations (notably for PFS, OS and TTNLT for patients treated with R-CHOP/R-CVP) were validated by experts.

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

1.6.1 Strengths

A good range of resources were searched and the searches were well documented making them transparent and reproducible. Supplementary searches of conference proceedings and HTA organisation websites were undertaken, along with a search of the ClinicalTrials.gov register in order to identify additional trials.

The company submission was largely in line with the NICE reference case. Utility scores were estimated using a mixed effects model based on observed EQ-5D data in the AUGMENT study.

The model was, in general, well-built and transparent. Apart from the base-case, the model provided ample opportunity for exploratory analyses using alternative assumptions on a range of input parameters.

1.6.2 Weaknesses and areas of uncertainty

The ERG was concerned about the overall quality of the searches conducted, as truncation and proximity operators were used inconsistently, and more synonyms could have been included in the search strategies. The date ranges of searches were not accurately reported. However, the searches were adequate, and given the range of resources searched, it was unlikely that any relevant studies were missed.

The results of the MAIC should be treated with a high degree of caution.

Similarly, the results of the economic evaluation should be treated with a high degree of caution, as the results of the MAIC serve as an important input parameter for the economic model. As the ERG did not have the necessary data to quantify uncertainty around the MAIC, model results do not include this structural uncertainty. Therefore, not only the company base-case, but also the ERG base-case and further exploratory analyses all (except those for the R-mono comparison) are conditional upon the possibly biased effectiveness of R^2 versus R-CHOP/R-CVP resulting from the MAIC. The ERG considers this to be a major source of uncertainty.

A main limitation was the lack of clarity and consistency in the selection of the parametric survival curves for extrapolation of PFS, OS and also TTNLT. The ERG considers particularly PFS to be overestimated for R^2 and underestimated (in the first year) for R-CHOP and R-CVP. Curve selection was often based only on avoiding implausible curve crossings, which may be indicative of a structural issue in the model design. For reference, the ERG would have liked to see the results of a state transition model next to the current partitioned survival model, but the company did not provide this.

Given the large impact of the FL-only scenario on the ICERs of the R-CHOP and R-CVP comparisons, the ERG considers the pooling of MZL and FL populations throughout the analysis to be another substantial source of uncertainty.

Lastly, the utility scores used in the model do not seem representative of the patient population. The ERG considers utilities in both progression free and progressed health states to be an overestimate.

1.7 Summary of exploratory and sensitivity analyses undertaken by the ERG

The ERG made various adjustments to the company's base-case, including the fixing of errors, violations, and amending the model according to the company's base-case according to its preferred assumptions (matters of judgement).

Fixing errors

1. Error cells when using 'van Oers' as input for R-CHOP efficacy

Fixing violations

2. Include AEs related to subsequent treatments in R^2 arm
3. Use pooled R-CVP/R-CHOP subsequent treatment rates instead of R-chemo. (Not applicable in the R-mono comparison)
4. Cap utilities at the general population level

Matters of judgment

5. Use exponential distribution to extrapolate OS in both arms
6. Use log-logistic for PFS in R^2 and Weibull for PFS in the comparator (not applied to R-mono comparison)
7. Used log-logistic for TTNLT both arms (not applied to R-mono comparison)

1.7.1 ERG probabilistic base-case results

The probabilistic ERG base-case ICER of R^2 versus R-CHOP was £15,818 per QALY gained (based on 1,000 iterations). This was slightly higher than the deterministic base-case ICER of £15,505. For R^2 versus R-CVP, the probabilistic ICER was £23,367 (deterministic £21,759) and for R^2 versus R-mono it was £29,010 (deterministic £27,372) (See Table 1.1). These rather substantial differences between probabilistic and deterministic ICERs were also observed in the company analyses (to a larger extent even).

Table 1.1: ERG probabilistic base-case results

Technologies	Total costs	Total QALYs	Incremental costs	Incremental QALYs	ICER (£/QALY)
Probabilistic ERG base-case for R^2 versus R-CHOP					
R^2	[redacted]	[redacted]	[redacted]	[redacted]	£15,818
R-CHOP	[redacted]	[redacted]			
Probabilistic ERG base-case for R^2 versus R-CVP					
R^2	[redacted]	[redacted]	[redacted]	[redacted]	£23,367
R-CVP	[redacted]	[redacted]			
Probabilistic ERG base-case for R^2 versus R-mono					
R^2	[redacted]	[redacted]	[redacted]	[redacted]	£29,010
R-mono	[redacted]	[redacted]			

ERG = Evidence Review Group = ICER = incremental cost effectiveness ratio; QALY = quality-adjusted life year

Deterministic scenario analyses were performed to examine the potential impact of alternative assumptions on the cost effectiveness estimates. For the R-CHOP/R-CVP comparisons, using R-CHOP and R-CVP efficacy from van Oers et al. would lower the ICER substantially, £8,251 for R^2 versus R-CHOP and £13,315 for R^2 versus R-CVP. Alternative assumptions regarding lowered utilities in the PP health states had the most significant upward impact, increasing the ICER to £33,626 for R^2 versus R-CHOP and £47,281 for R^2 versus R-CVP. For the R-mono comparison, lowering the PP health state utility had the opposite effect, lowering the ICER to £17,826. Another influential scenario was the change of time-point where treatment waning starts to three years (instead of five years in base-case). This increased the ICER to £40,543.

In conclusion, even though the ERG base-case ICER for R-CHOP was below £20,000, the uncertainty around the cost effectiveness of R^2 is substantial, mainly caused by the possible bias introduced by the indirect treatment comparison, which could not be accounted for in the ERG analyses. The ICER for R-CVP is higher and suffers from the same uncertainty. The R-mono analysis is based on a direct comparison, but is also surrounded by substantial uncertainty, as the ICER is rather sensitive to, for instance, the time-point at which treatment waning starts and utilities in the PP health state.

2. BACKGROUND

In this report, the ERG provides a review of the evidence submitted by Celgene in support of lenalidomide (Revlimid®) in combination with rituximab (MabThera®) (R2), for the treatment of adults with treated follicular lymphoma (FL) or marginal zone lymphoma (MZL).

We will outline and critique the company's description of the underlying health problem and the overview of current service provision. The information is taken from section B.1.3 of the company's submission (CS) with sections referenced as appropriate. For additional information on the aetiology, epidemiology, health impact, prognosis and management of FL or MZL, please see the CS (pages 13-23).[1]

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

The underlying health problems in this appraisal are follicular lymphoma (FL) and marginal zone lymphoma (MZL), the two most common subtypes of indolent non-Hodgkin lymphoma (iNHL).

As described in the CS, FL is typically characterised by an indolent clinical course, with recurrent remissions and relapses; with each relapse, the disease becomes more resistant and/or refractory to treatment and each remission becomes shorter than the preceding one.[1] The incidence of FL increases with age, with a median presentation between 60 and 65 years, and a slightly higher incidence in females.[2] At diagnosis, most patients have advanced disease (Stage III: 18.4%; Stage IV: 50.5%).[3] The overall five-year relative survival rate for patients with FL in the UK is 89% and specifically for Stages III and IV, is approximately 80%.[4, 5] Since the introduction of rituximab, the median OS of patients with FL has extended to 20 years in some studies,[6] compared with nine years previously reported.[7] Despite the available treatment options, most patients eventually die from this disease.[8]

Patients with MZL represent a generally older (median age at diagnosis is 70–73 years)[2] and more advanced population compared with those with FL.[2, 3, 9] The primary organ of origin is the most significant prognostic factor and dictates organ-specific management strategies.[9] Patients with MZL have a similar prognosis to those with FL. In the UK, the overall five-year survival ranges between 77% and 90% depending on the subtype of MZL.[4] The median OS for UK patients has been reported as between eight and 12.6 years, depending on the subtype of MZL.[10, 11]

For FL, the CS notes that the Office of National Statistics (ONS) estimates 2,168 patients were diagnosed with FL in 2017 in England.[12] Of these, [redacted]% (n=[redacted]) have first-line chemotherapy, while [redacted]% (n=[redacted]) undergo a 'watch and wait' approach,[3] of which [redacted]% (n=[redacted]) go on to receive chemotherapy.[13] Therefore, the total number of FL patients on first-line chemotherapy is [redacted]. Of these, [redacted]% (n=[redacted]) are expected to receive second-line chemotherapy or beyond.[13] For MZL, the CS states that based on the anticipated figures for the different MZL types, the total number of MZL patients in England in 2017 is estimated at 1,411.[14-16] Of these, 34.9% (n=492) have first-line chemotherapy, while 49.9% (n=704) undergo a 'watch and wait' approach[3] of which [redacted]% (n=[redacted]) go on to receive chemotherapy.[13] Therefore, the total number of MZL patients on first-line chemotherapy is [redacted]. Of these, [redacted]% (n=[redacted]) are expected to receive second-line chemotherapy or beyond.[13] The ERG has no reason to doubt these numbers.

2.2 Critique of company's overview of current service provision

In the CS, lenalidomide is described as an agent that binds to cereblon in the Cullin-4 RING E3 ubiquitin ligase that promotes the degradation of the haematopoietic transcription factors Ikaros and Aiolos.[17, 18] As a result lenalidomide inhibits proliferation and enhances apoptosis of certain haematopoietic tumour cells (including FL and MZL tumour cells), enhances T cells and natural killer (NK) cell-mediated

immunity and increases the number of NK, T and NK T cells. Single agent lenalidomide reactivates dysfunctional T and NK cells from FL patients.[17] Rituximab is an anti-CD20 antibody; its mechanisms of action are to augment NK cell-mediated killing of malignant B cells via antibody-dependent cellular cytotoxicity (ADCC), to enhance antibody-dependent cellular phagocytosis (ADCP) and to induce complement-mediated killing.[17] The combination immunotherapy of lenalidomide and rituximab acts by complementary mechanisms including direct tumour apoptosis in FL and MZL and immune-mediated activities, such as activation of NK cells and immune synapse formation, resulting in increased ADCC in vitro.[18]

The CS describes the following sources that were used in the company's interpretation of the positioning of R^2 in the treatment pathway for FL (see Figure 2.1): an advisory board conducted by Celgene in March 2019, involving six UK clinical experts in NHL and two health economics experts,[19] ad-hoc follow up with advisors, technology appraisal 472 (TA472, Obinutuzumab with bendamustine for treating follicular lymphoma refractory to rituximab),[20] TA243 (Rituximab for the first-line treatment of stage III-IV follicular lymphoma),[21] TA226 (Rituximab for the first-line maintenance treatment of follicular non-Hodgkin's lymphoma),[22] TA513 (Obinutuzumab for untreated advanced follicular lymphoma),[23] NICE treatment pathway for FL[24] and NICE guideline for the diagnosis and management of non-Hodgkin's lymphoma (NG52).[25]

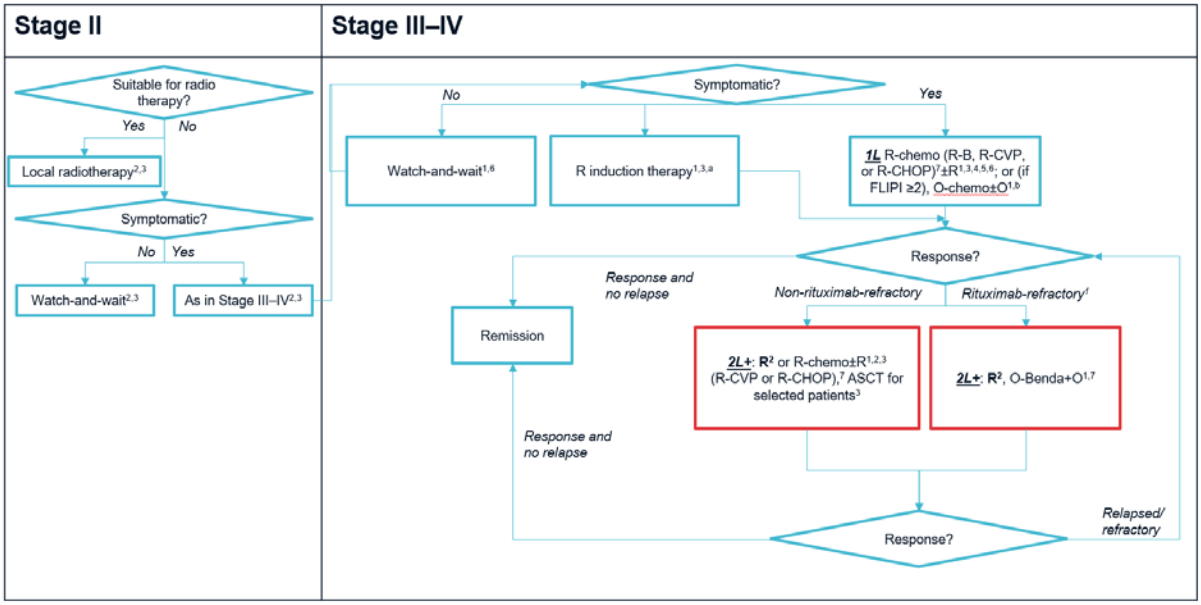
Figure 2.1 shows the treatment pathway proposed by the company for patients with follicular lymphoma. The flowchart distinguishes between Stage II and Stages III and IV. For Stage II FL radiotherapy is advised as a first-line option when suitable, when radiotherapy is unsuitable 'watch and wait' should be the preferred approach for asymptomatic patients, while symptomatic patients should be treated as in Stages III and IV. For asymptomatic patients with advanced Stages III and IV a 'watch and wait' approach or rituximab induction therapy are recommended in first-line. If patients present with symptoms, pharmacological therapy is recommended in first-line (i.e. rituximab-chemotherapy (R-bendamustine, R-CVP or R-CHOP) with or without rituximab maintenance therapy). For patients with a Follicular Lymphoma International Prognostic Index (FLIPI) score of 2 or more obinutuzumab-chemotherapy with or without obinutuzumab maintenance therapy may be given as first-line therapy.

Second-line therapy depends on whether patients are refractory to rituximab or not, according to the company's proposed pathway. The ERG has questioned this in the clarification letter (Clarification letter, Question A.7).[26] To the ERG it seems counter intuitive that rituximab containing treatments are not appropriate for rituximab-refractory patients, but rituximab in combination with lenalidomide is. The company stated that this was done to 'reflect the current approach to patient management in the UK' and that 'patients determined to be R-refractory are treated differently to the non-R-refractory population in the UK due to the availability of an alternative treatment, obinutuzumab-bendamustine, approved in the EU and recommended by NICE' (Response to Clarification Letter).[26] However, obinutuzumab in combination with bendamustine is only used as part of the Cancer Drugs Fund; this means 'there is significant remaining clinical uncertainty which needs more investigation, through data collection in the NHS or clinical studies'.[27] This also means that it is not considered a relevant comparator for disease that is refractory to rituximab by NICE.[28]

Depending on the response to first-line therapy, patients who are not refractory to rituximab may be given rituximab-chemotherapy (R-CVP or R-CHOP) with or without rituximab as maintenance. Autologous stem cell transplant (ASCT) may be an option for selected patients at this stage. Patients who were refractory to rituximab are recommended obinutuzumab-bendamustine (O-Benda) with obinutuzumab as maintenance. Rituximab in combination with lenalidomide (R^2) is an option for both, R-refractory patients and non-R-refractory patients in second-line.

As justification for not including rituximab monotherapy as an option for non-R-refractory patients in second-line, the company cited the opinion of clinical experts, elicited during the advisory board meeting conducted by Celgene in March 2019:[19] *"According to clinical experts, R mono is rarely used in the relapsed/refractory setting in UK clinical practice."* Clinical experts also advised that: *"R-Benda is primarily used in a first-line setting and clinicians are reluctant to re-challenge relapsed/refractory patients with bendamustine in subsequent lines of therapy."* Therefore, bendamustine monotherapy was not considered an option in the R-refractory population.

Figure 2.1: Treatment pathway as described by the company for patients with follicular lymphoma with proposed positioning of R²



Source: Section B.1.3 of the CS.[1]

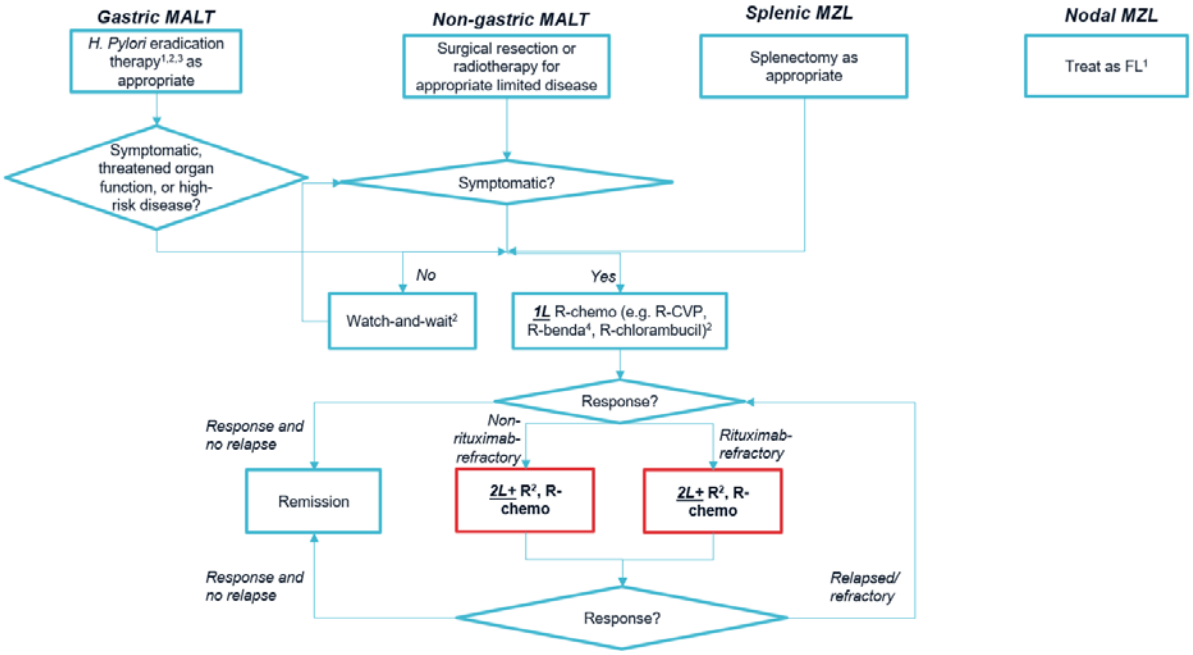
1L = first-line; 2L = second-line; Benda = bendamustine; CHOP = cyclophosphamide, doxorubicin, vincristine, prednisolone; CVP = cyclophosphamide, vincristine, prednisolone; O = obinutuzumab; R = rituximab; R-chemo+R = rituximab and chemotherapy induction followed by rituximab maintenance therapy; ASCT = autologous stem cell transplant.

* Please note that references in the graph are references from the CS.

The CS describes the following sources that were used in the company's interpretation of the positioning of R² in the treatment pathway for MZL (see Figure 2.2): an advisory board conducted by Celgene in March 2019, involving six UK clinical experts in NHL and two health economics experts,[19] NICE guideline for the diagnosis and management of non-Hodgkin's lymphoma (NG52),[25] the ESMO guidelines for marginal zone lymphoma, mantle cell lymphoma and peripheral T-cell lymphoma,[29] and the fact sheet for MZL by the Lymphoma Research Foundation.[30]

Figure 2.2 shows the treatment pathway proposed by the company for patients with marginal zone lymphoma (MZL). Treatment options are dependent on the type of MZL: gastric or non-gastric mucosa associated lymphoid tissue (MALT), splenic marginal zone lymphoma (SMZL) or nodal marginal zone lymphoma (NMZL). First-line treatment options include R-chemo (e.g. R-CVP, R-Benda or R-chlorambucil). Second-line treatment options are R² or R-chemo, both for R-refractory patients and for non-R-refractory patients. It is not clear to the ERG why R-chemo is a second-line treatment option for R-refractory patients with MZL, but not for R-refractory patients with FL.

Figure 2.2: Treatment pathway as described by the company for patients with marginal zone lymphoma with proposed positioning of R[2]



Source: Section B.1.3 of the CS.[1]

Benda = bendamustine; CHOP = cyclophosphamide, doxorubicin, vincristine, prednisolone; CVP = cyclophosphamide, vincristine, prednisolone; FL = follicular lymphoma; H = Helicobacter; MALT = mucosa-associated lymphoid tissue; MZL = marginal zone lymphoma; R = rituximab.

References: 1. Dreyling 2013;[29] 2. NICE, 2016;[25] 3. Lymphoma Research Foundation, 2018;[30] 4. Celgene, 2019.[19]

3. CRITIQUE OF COMPANY'S DEFINITION OF DECISION PROBLEM

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

	Final scope issued by NICE	Decision problem addressed in the company submission	Rationale if different from the final NICE scope	ERG Comment
Population	Adults with treated follicular lymphoma or marginal zone lymphoma	Adults with treated follicular lymphoma or marginal zone lymphoma	N/A	The population is in line with the scope. However, R^2 does not currently have a UK marketing authorisation, although CHMP opinion is anticipated on [redacted], and marketing authorisation is expected in [redacted].
Intervention	Lenalidomide with rituximab (R^2)	Lenalidomide with rituximab (R^2)	N/A	The intervention is in line with the scope.
Comparator(s)	• Rituximab monotherapy (R-mono) • Rituximab in combination with chemotherapy • Established clinical management without lenalidomide (including but not limited to bendamustine)	For non-rituximab refractory patients: • Rituximab in combination with chemotherapy • Rituximab plus cyclophosphamide, doxorubicin, vincristine, prednisolone (R-CHOP) • Rituximab plus cyclophosphamide vincristine prednisolone (R-CVP) For rituximab refractory patients: • Established clinical management without lenalidomide	For non-rituximab refractory patients: • R-mono is not considered a relevant comparator as clinical expert opinion confirmed it is rarely used in the relapsed/refractory setting in the UK.[19, 31] For rituximab refractory patients: • O-Benda is included as an option for rituximab-refractory patients under the category 'Established clinical management without lenalidomide'. This is the only	

	Final scope issued by NICE	Decision problem addressed in the company submission	Rationale if different from the final NICE scope	ERG Comment
		o Obinutuzumab in combination with bendamustine (O-Benda)	NICE-recommended option for this patient group (via the CDF) and clinical experts stated this is the likely treatment choice for FL patients refractory to rituximab.[19] • Bendamustine monotherapy (Benda mono) is not considered a comparator in this population given that clinical experts believe O-Benda has largely replaced use of Benda mono in rituximab refractory patients.[19]	
Outcomes	The outcome measures to be considered include: • Overall survival • Progression-free survival • Overall response rate • Adverse effects of treatment • Health-related quality of life	The outcome measures to be considered include: • Overall survival • Progression-free survival • Event-free survival • Overall response rate • Adverse effects of treatment • Health-related quality of life • Time to next anti-lymphoma treatment • Time to next chemotherapy treatment • Response rate to next anti-lymphoma treatment	Several efficacy outcomes have been presented in addition to those in the scope as several secondary and exploratory outcomes were reported in the AUGMENT and MAGNIFY studies that provide additional insight into the efficacy of R^2	All outcomes are reported in AUGMENT. However, for the indirect comparisons only a limited number of outcomes have been included.

	Final scope issued by NICE	Decision problem addressed in the company submission	Rationale if different from the final NICE scope	ERG Comment
Economic analysis	The reference case stipulates that the cost effectiveness of treatments should be expressed in terms of incremental cost per quality-adjusted life year. The reference case stipulates that the time horizon for estimating clinical and cost effectiveness should be sufficiently long to reflect any differences in costs or outcomes between the technologies being compared. Costs will be considered from an NHS and Personal Social Services perspective. The availability of any PAS for the intervention or comparator technologies will be taken into account.	Adhering to the reference case, the cost effectiveness of treatments is expressed in terms of incremental cost per quality adjusted life year. Adhering to the reference case, a lifetime horizon is used. Adhering to the reference case the economic analyses has been conducted from an NHS and Personal Social Services perspective Adhering to the reference case, the PAS has been applied in all economic analysis for all Celgene products.	Confidential PAS schemes that apply to relevant subsequent comparator therapies are not included in these analyses as Celgene is not privy to such information	
Subgroups to be considered	None listed in scope	No specific subgroups	N/A	

Source: CS, Table 1, pages 7-9.

CDF = Cancer Drugs Fund; FL = follicular lymphoma; MZL = marginal zone lymphoma; NICE = National Institute for Health and Care Excellence.

3.1 Population

The population defined in the scope is: Adults with previously treated follicular lymphoma or marginal zone lymphoma.[32] The population in the CS is in line with the NICE scope.[1]

According to the company R[2] does not currently have a UK marketing authorisation, although CHMP opinion is anticipated on [redacted], and marketing authorisation is expected in [redacted]. Therefore, the relevant population for this appraisal is currently unclear.

The anticipated license is as follows: Lenalidomide in combination with rituximab (anti-CD20 antibody) is indicated for the treatment of adult patients with previously treated FL or MZL.[18] Treatment should not be initiated in patients with hypersensitivity to the active substance or to any of the excipients, in women who are pregnant, in women of childbearing potential unless all of the conditions of the Pregnancy Prevention Programme are met, and in children and adolescents from birth to less than 18 years.

3.2 Intervention

The intervention (lenalidomide in combination with rituximab) is in line with the scope.

Lenalidomide is administered orally and rituximab is administered by intravenous (IV) infusion. Lenalidomide capsules should be taken orally at about the same time on the scheduled days.[18] The recommended starting dose of lenalidomide is 20 mg, orally once daily on days 1 to 21 of repeated 28-day cycles for up to 12 cycles of treatment. The recommended starting dose of rituximab is 375 mg/m^2 IV every week in cycle 1 (days 1, 8, 15, and 22) and day 1 of every 28-day cycle for cycles 2 through 5.[18]

The following tests/investigations are recommended when administering lenalidomide in combination with rituximab:[18]

- Medically supervised pregnancy tests with a minimum sensitivity of 25 mIU/mL must be performed for women of childbearing potential, including those who practice abstinence, before treatment, every four weeks during treatment, and four weeks after the end of treatment (except in the case of confirmed tubal sterilisation)
- Patients with known risk factors for myocardial infarction (including prior thrombosis) should be closely monitored, and action should be taken to try to minimise all modifiable risk factors (e.g. smoking, hypertension, and hyperlipidaemia)
- A complete blood cell count should be performed at baseline and then weekly for the first three weeks of Cycle 1 (28 days), every two weeks during Cycles 2 through 4, and then at the start of each cycle thereafter
- Careful monitoring and evaluation for tumour flare reaction (TFR) is recommended.
- Careful monitoring and evaluation for tumour lysis syndrome (TLS) is recommended. Patients should be well hydrated and receive TLS prophylaxis, in addition to weekly chemistry panels during the first cycle or longer, as clinically indicated.

3.3 Comparators

The description of the comparators in the NICE scope is as follows: Rituximab monotherapy, rituximab in combination with chemotherapy, and established clinical management without lenalidomide (including but not limited to bendamustine).[32]

ERG comment: The NICE scope does not make a distinction in terms of patients being rituximab refractory or not. However, the CS has different comparators for rituximab refractory patients and non-rituximab refractory patients. The company was asked why they made this distinction (Clarification Letter, Question A7), because, according to the ERG, if the intervention includes rituximab, the comparator should also be able to include rituximab. The company stated that this was done to 'reflect the current approach to patient management in the UK' and that 'patients determined to be R-refractory are treated differently to the non-R-refractory population in the UK due to the availability of an alternative treatment, obinutuzumab-bendamustine, approved in the EU and recommended by NICE' (Response to Clarification Letter).[26]

For non-rituximab refractory patients, the company included two comparators, both different types of rituximab in combination with chemotherapy:

- Rituximab plus cyclophosphamide, doxorubicin, vincristine, prednisolone (R-CHOP)
- Rituximab plus cyclophosphamide vincristine prednisolone (R-CVP)

For rituximab refractory patients, the company included one comparator:

- Established clinical management without lenalidomide, i.e. obinutuzumab in combination with bendamustine (O-Benda).

ERG comment: The ERG has several concerns with these comparators. Firstly, the NICE scope does not make a distinction in terms of patients being rituximab refractory or not. Therefore, the CS should have included a comparison of R^2 with rituximab monotherapy for all patients as specified in the scope. Secondly, even if the NICE committee accepts splitting the population in rituximab refractory patients and non-rituximab refractory patients, the CS should still have included a comparison with rituximab monotherapy for both populations as specified in the scope. Thirdly, the company included O-Benda as a comparator for rituximab refractory patients. However, in the response from NICE to comments on the draft scope, NICE clearly stated that "obinutuzumab in combination with bendamustine is only used as part of the Cancer Drugs Fund therefore it is not considered a relevant comparator for disease that is refractory to rituximab." (see NICE Response to comments on draft scope (page 4).[28] Therefore, we believe that the submission currently does not present any relevant evidence for R-refractory patients.

3.4 Outcomes

The NICE final scope lists the following outcome measures:

- overall survival
- progression-free survival
- overall response rate
- adverse effects of treatment
- health-related quality of life.

These outcomes were all assessed in the AUGMENT trial. The company included several additional outcomes (event-free survival, time to next anti-lymphoma treatment, time to next chemotherapy treatment, and response rate to next anti-lymphoma treatment) based on the AUGMENT trial. Therefore, all these outcomes are available for the comparison R^2 versus rituximab monotherapy.

All other comparisons rely on indirect comparisons. The company was not able to find any evidence providing a common comparator linking R^2 with any of the comparators of interest (apart from rituximab monotherapy, which was dismissed by the company). Therefore, the company performed a matching-adjusted indirect comparison (MAIC) to compare R^2 with R-CHOP and R-CVP in the non-

R-refractory population. For these analyses, only the outcomes OS, PFS, overall response rate (ORR) and complete response (CR) rate were used.

3.5 Other relevant factors

According to the company, R^2 represents an innovation in the management of patients with previously treated FL and MZL, because it is the first chemotherapy-free combination immunotherapy regimen licensed in this setting by the US Food and Drug Administration. The regimen is currently pending approval in the EU (CS, Document A, Section A16, pages 36-37; and Document B, Section B.2.12, pages 98-99).[1, 33]

There is a confidential simple discount PAS for lenalidomide ([redacted]) which applies to all current and future indications.

End-of-life criteria are not applicable for this appraisal (see CS, page 105).[1]

According to the company, no equality considerations have been identified or are anticipated (see CS, Document A, Section A3, page 8; and Document B, Section B.1.4, page 23).[1, 33]

4. CLINICAL EFFECTIVENESS

4.1 Critique of the methods of review(s)

4.1.1 Searches

Appendix D.1.1 of the CS provided details of the systematic search of the literature used to identify clinical effectiveness literature. It was reported that searches were conducted on 1 September 2017 and then updated on 4 April 2019. The ERG clarification letter asked whether the update searches had been conducted from database inception. In response, the company stated that it had been incorrect to report that the searches had been updated from September 2017 to 4 April 2019; the searches conducted on 4 April 2019 replaced the 2017 searches. "This was a de novo Clinical SLR conducted to replace the older Clinical SLR (with a cut off of September 2017), as some changes were made to the protocol and search strategies were made more extensive. All searches were conducted from database inception."[26] A summary of the resources searched is provided in Table 4.1.

Table 4.1: Resources for the clinical effectiveness and adverse reactions literature searches

Search strategy element	Resource	Host/Source	Date Range	Date searched
Electronic databases	MEDLINE	ProQuest	Not reported	4 April 2019
	Embase		Not reported	
	CENTRAL	Cochrane Library	Not reported	
	CDSR		Not reported	4 April 2019
Conference proceedings	EHA	Organisation websites, abstract books	2015-18	April 2019
	ICML		2013, 2015, 2017	
	ASCO		2015-2018	
	ASH		2014-2018	
	ESMO		2014-2018	
HTA Agencies	NICE	Organisation websites		April 2019
	CADTH			
	TGA			
Trials registries	ClinicalTrials.gov	ClinicalTrials.gov		April 2019
Manual searching of references of published systematic reviews, meta-analyses, and HTA documents was also conducted to identify potential publications that may not have been identified from the electronic searches.				
CENTRAL = Cochrane Central Register of Controlled Trials; CDSR = Cochrane Database Systematic Reviews; EHA = European Hematology Association; ICML = International Conference on Malignant Lymphomas; ASCO = American Society of Clinical Oncology; ASH = American Society of Hematology; ESMO = European Society for Medical Oncology; NICE = National Institute for Health and Care Excellence; CADTH = Canadian Agency for Drugs and Technologies in Health; TGA = Therapeutic Goods Administration Australia.				

ERG comment:

- The selection of databases searched was adequate, and searches were clearly reported and reproducible. The database name, host and date searched were provided. The date range of the searches was not reported.
- Searches conducted in September 2017 were reported in the CS, along with 'update' searches conducted in April 2019. In response to the ERG clarification letter, the company explained that

the April 2019 searches had replaced the September 2017 searches. Reporting the April 2019 searches would have been sufficient.

- An extensive range of resources additional to database searches was included in the SLR to identify further relevant studies and grey literature. Details of the resources searched, search strategies or search terms used, dates of searches, and results were not reported in the CS, but full details of the conference proceedings and HTA organisation website searches were provided in response to the ERG clarification letter.
- Accurate details of the MEDLINE segments searched were not reported. It is not clear if MEDLINE In-Process, Ahead of Print, and Daily Update were searched.
- Truncation and proximity operators were inconsistently used throughout. There were few synonyms used in the 2017 searches, and although there were more included in the 2019 'update' searches, they were still lacking.
- Comparators of interest were not included in the 2017 searches, but were included in the 2019 searches: prednisolone and cyclophosphamide.
- As study design filters were not included, both efficacy and safety evidence could be identified.
- The Cochrane Library searches did not report the database issue searched.
- The CS reported that ClinicalTrials.gov was searched for trials, but limited to "studies with results". In response to the ERG clarification letter, the company supplied full details of the ClinicalTrials.gov searches conducted in April 2019, which searched for "all studies".
- The PRISMA flow diagram (Figure 1) provided in the CS suggested that the 'update' searches of April 2019 were conducted from database inception, and replaced the original September 2017 search results. This was confirmed in the company response to clarification.
- A good range of conference proceedings and HTA organisation websites were searched, and although full details of these searches were not provided in the CS, they were provided in response to the ERG clarification letter.

4.1.2 Inclusion criteria

The eligibility criteria used in the search strategy for randomised controlled trials (RCTs) and non-RCTs is presented in Table 4.2.

Table 4.2: Eligibility criteria (PICOS scope)

	Inclusion criteria	Exclusion criteria
Population (P)[a]	• Adults (≥18+) with relapsed and/or refractory FL • Adults (≥18+) with relapsed and/or refractory MZL • Any stage of disease	• Patients <18 years of age • Patients that do not have R/R FL/MZL
Intervention (I)	*Systemic induction (i.e., chemo-therapy, immunotherapy, chemo-immunotherapy) therapies recommended by NCCN/ESMO and deemed relevant to current clinical practice:* • Rituximab + bendamustine [R-B] • Rituximab + lenalidomide [R^2] • Rituximab + cyclophosphamide + vincristine + prednisone [R-CVP] • Rituximab + cyclophosphamide + doxorubicin + vincristine + prednisone [R-CHOP]	Any treatments that are not listed under the inclusion criteria

	Inclusion criteria	Exclusion criteria
	• Rituximab + chlorambucil [R-Chl] • Obinutuzumab + bendamustine [O-B] • Obinutuzumab + lenalidomide • Rituximab alone • Bendamustine alone • Lenalidomide alone • Idelalisib • Ibrutinib • Copanlisib • Tazemetostat • Rituximab + mitoxantrone + Chlorambucil + prednisone (R-MCP) • Rituximab + cyclophosphamide + doxorubicin + etoposide+ prednisone + interferon alpha (R-CHVPI)	
Comparators (C)[b]	• Any of the interventions listed in inclusion criteria OR fludarabine containing regimen • Placebo	Any treatments that are not listed under the inclusion criteria
Outcomes	• Survival (overall, progression-free, disease-free) • Response (overall, complete, partial) • Duration of treatment (median) • Duration of response • Quality of life: EORTC-QLQ-C30, EQ-5D, FACT-g and FACT-lym • Time to next lymphoma treatment • Adverse events of interest	Outcomes not included under inclusion criteria
Study design (S)	• Randomised controlled trials (RCTs) • Non-randomised clinical trials • Observational cohort studies (retrospective or prospective) • Systematic reviews and meta-analyses (for identification of primary studies only) • Single arms studies • Cross-sectional studies, case-control studies • Comparative studies	• Case series/case reports • Studies of non-original data • Non-systematic reviews • Comment, editorial, letter • Theses and dissertations • Non-human studies • Pharmacokinetic, pharmacodynamic, and bioequivalence studies
Publication type	Sample size ≥20 participants meeting the target population[c]	Sample size <20 participants meeting the target population[b]
Language	English language	Non-English

Source: CS, Appendix D1, Table 7.[1]
FL = follicular lymphoma; MZL = marginal zone lymphoma.
Notes: a) ≥70% of a mixed population needs to have R/R FL/MZL, or results need to be reported as subgroup data for the patient population of interest; b) only applicable to comparative studies; c) sample size limitation

	Inclusion criteria	Exclusion criteria
applies only to non-randomised studies. RCTs will be included regardless of sample size.		

ERG comment: Generally, the inclusion criteria are in line with the NICE scope. There are two small issues, both relating to outcomes. First, looking at inclusion criteria as formulated in the CS, it seems that only four specific quality of life instruments (EORTC-QLQ-C30, EQ-5D, FACT-g and FACT-lym) were included. Therefore, a paper comparing R^2 with R-chemo reporting the results for the SF-36 would be excluded. Second, only studies that reported 'adverse events of interest' were included. However, it is not specified what 'adverse events of interest' are. According to the ERG, all quality of life instruments and all adverse events should be eligible for inclusion. Nine studies were excluded because they did not include any relevant outcomes (CS, Appendix D, Figure 1, page 17).[34] However, the company did not provide a list with references of these studies; therefore, the ERG are unable to check whether any of these studies might be relevant.

4.1.3 Critique of data extraction

Data extraction of the selected relevant studies for the clinical evidence was performed by two independent reviewers and any discrepancies between reviewers were resolved by consensus and/or in conjunction with a third reviewer. The CS explains that when multiple sources of the same data were reported all sources were reviewed and reconciled (CS, Appendix D, page 15).[34]

ERG comment: The process of data extraction appears well conducted. The extraction by two independent reviewers minimises the risk of error and bias.

4.1.4 Quality assessment

In section D.5 of Appendix D of the CS,[34] the company lists the signalling questions that supported the risk of bias assessment of the trials AUGMENT and MAGNIFY, as follows:

- Was randomisation carried out appropriately?
- Was the concealment of treatment allocation adequate?
- Were the groups similar at the outset of the study in terms of prognostic factors, for example, severity of disease?
- Were the care providers, participants and outcome assessors blind to treatment allocation? If any of these people were not blinded, what might be the likely impact on the risk of bias (for each outcome)?
- Were there any unexpected imbalances in drop-outs between groups? If so, were they explained or adjusted for?
- Is there any evidence to suggest that the authors measured more outcomes than they reported?
- Did the analysis include an intention-to-treat analysis? If so, was this appropriate and were appropriate methods used to account for missing data?

In the final statement regarding the quality assessment of the AUGMENT trial, the CS reports that '*Subsequently, this double-blind randomization method ensured low levels of bias in the AUGMENT study*'.[1] With regard to the to the quality assessment of the MAGNIFY trial, the CS states that *'(…) therefore, a lack of blinding was not thought to have a considerable effect on the outcome of the study. Furthermore, the results of interest for this submission are taken from the initial treatment period only and are therefore not affected by the open-label design'*.[1]

ERG comment: It is recommended that two reviewers perform risk of bias/quality assessment independently of each other to reduce the potential for any errors. This is not described in the CS.

Regarding the quality assessment of the AUGMENT trial, the ERG agrees that this is a good quality double-blind randomised trial. Regarding the MAGNIFY trial, the company only used data from the induction phase of the trial, i.e. before randomisation. Therefore, this study should be assessed as a single arm study, not an RCT. As such, the single arm from the MAGNIFY study is at high risk of bias.

4.1.5 Evidence synthesis

The company did not perform a meta-analysis to pool the two R^2 studies, AUGMENT and MAGNIFY.

ERG comment: The ERG agrees that this is justified because the MAGNIFY study, as used in the CS, did not have a comparator arm; and because there are important differences between the populations in the two studies. In particular MAGNIFY included both R-refractory and non-refractory patients but AUGMENT was only non-refractory patients, and there were differences regarding age, previous rituximab, refractory to last regimen, line of therapy, disease stage and Eastern Cooperative Oncology Group (ECOG) performance status.

The company did perform indirect comparisons because, according to the company, 'No head-to-head data are available for R^2 versus any of the comparators of interest to this submission; only R-mono was compared with R^2 within the AUGMENT RCT' (CS, Section b.2.9, page 67).[1] The ERG disagrees with this statement because, according to the NICE scope,[32] rituximab monotherapy is a relevant comparator; therefore, there are relevant head-to-head data available.

The company performed two matching-adjusted indirect comparisons (MAIC), one for the rituximab refractory population and one for the non-rituximab refractory population.

For non-rituximab refractory patients, the company included two comparators, both different types of rituximab in combination with chemotherapy:

- Rituximab plus cyclophosphamide, doxorubicin, vincristine, prednisolone (R-CHOP)
- Rituximab plus cyclophosphamide vincristine prednisolone (R-CVP)

For rituximab refractory patients, the company included one comparator:

- Established clinical management without lenalidomide, i.e. obinutuzumab in combination with bendamustine (O-Benda).

ERG comment: The ERG has several concerns with these comparators. Firstly, the NICE scope does not make a distinction in terms of patients being rituximab refractory or not. Therefore, the CS should have included a comparison of R^2 with rituximab monotherapy for all patients as specified in the scope. Secondly, even if the NICE committee accepts splitting the population in rituximab refractory patients and non-rituximab refractory patients, the CS should still have included a comparison with rituximab monotherapy for both populations as specified in the scope. Thirdly, the company included O-Benda as a comparator for rituximab refractory patients. However, in the response from NICE to comments on the draft scope, NICE clearly stated that "obinutuzumab in combination with bendamustine is only used as part of the Cancer Drugs Fund therefore it is not considered a relevant comparator for disease that is refractory to rituximab." (see NICE Response to comments on draft scope (page 4)[28]). Therefore, the ERG believes that the submission currently does not present any relevant evidence for R-refractory patients.

Methods and results of the indirect comparison for the non-rituximab refractory population, R^2 versus R-CHOP and R-CVP, are discussed in Section 4.4 of this report.

Methods and results of the indirect comparison for the rituximab refractory population, R^2 versus O-Benda, will be ignored as this is not a relevant comparator according to NICE.[28]

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

4.2.1 Included studies

The company identified three randomised controlled trials (RCTs) of the intervention of interest (lenalidomide in combination with rituximab, R^2): the AUGMENT trial,[35] the MAGNIFY trial,[36] and the ALLIANCE trial;[37] and one non-RCT: Tuscano 2014.[38] In this ERG report, the focus will be on the AUGMENT trial,[35] because this provides a head-to-head comparison of the intervention of interest (lenalidomide in combination with rituximab, R^2) versus a relevant comparator according to the NICE scope (rituximab monotherapy).

The other three studies of the intervention of interest (R^2) will be ignored in this report for the following reasons:

- The ALLIANCE trial[37] is a randomised, multicentre, Phase II study of R^2 versus lenalidomide monotherapy in patients with previously treated FL and prior rituximab. Lenalidomide monotherapy is not a relevant comparator according to the NICE scope. Therefore, only one arm of this trial is relevant.
- The MAGNIFY trial[36] is an ongoing, randomised, open-label, multicentre, Phase IIIb study of R^2 induction therapy followed by either R^2 maintenance therapy or R-mono maintenance therapy in patients with FL Grade 1–3b, MZL, or mantle cell lymphoma. Only patients who had stable disease (SD), partial response (PR), complete response (CR) or complete response unconfirmed (CRu) at the end of 12 cycles of initial therapy were randomised 1:1 to receive R^2 maintenance therapy or rituximab maintenance therapy. In the CS, the company only used data from the induction phase (before randomisation). However, this is single arm data, and as there is relevant RCT data from the AUGMENT trial, these data will be ignored in this report.
- Tuscano 2014[38] is a single-arm Phase II study evaluating the safety and efficacy of lenalidomide in combination with rituximab in patients with relapsed/refractory, indolent non-Hodgkin lymphoma (NHL), including 30 patients (22 FL, three MZL and five other NHL).

4.2.2 Methodology of the AUGMENT trial

The AUGMENT trial[35] is a randomised, double-blind, multicentre, Phase III study of R^2 versus rituximab plus placebo (R-mono) in non-R-refractory patients with FL Grade 1–3a or MZL. The study was conducted across 96 sites in 17 countries. The number of sites and patients from the UK have not been reported in the CS; but according to the clinical study report (CSR), the trial did not include any patients from the UK.

To be eligible for inclusion in the study, patients had to be aged ≥18 years, with histologically confirmed MZL or Grade 1, 2, or 3a FL (Grade 3b FL patients were excluded). Patients were required to have been previously treated with at least one systemic chemotherapy, immunotherapy or R-chemo. Initially, rituximab-naïve patients were included in the study; however, a protocol change required patients to have received at least two previous doses of rituximab. This change was carried out to ensure a study population that aligned with a population commonly seen in clinical practice. Furthermore, patients had to have documented relapsed/refractory FL or MZL; however, R-refractory patients were excluded (full inclusion and exclusion criteria are presented in Appendix M1 of the CS,[34] a summary is presented below in Table 4.3).

During the treatment period, patients underwent efficacy and safety assessments for a maximum of 12 cycles. Patients received oral lenalidomide or placebo at a starting dose of 10 mg (if creatine clearance (CrCl) ≥30 mL/min and <60 mL/min) or 20 mg (if CrCl ≥60 mL/min) once daily on Days 1 to 21 in each 28-day cycle, combined with four-weekly infusions of rituximab intravenously (IV) at a dose of 375 mg/m^2, followed by four additional doses on Day 1 of Cycles 2, 3, 4, and 5. Patients were stratified by prior rituximab treatment (yes vs. no), time since last anti-lymphoma therapy (≤2 vs. >2 years), and histology (FL vs. MZL), and then randomised 1:1 to R^2 or R-mono for 12 cycles. Treatment was terminated upon relapse or progression of disease, withdrawal of consent, or unacceptable toxicity.

Primary efficacy analyses were conducted on the ITT population, defined as all randomised patients. The primary endpoint of the study was PFS, as assessed by the Independent Review Committee (IRC) using a modification of the 2007 International Working Group Response Criteria (IWGRC [i.e. without a positron emission tomography scan]). Efficacy was assessed further in the ITT population through a number of secondary endpoints, including overall response rate (ORR), complete response (CR) rate, time to next anti-lymphoma treatment (TTNLT), duration of response (DOR), durable complete response rate (DCRR; defined as the proportion of patients that stayed in complete response for at least one year) and duration of complete response (DOCR). Safety analyses were conducted on the safety population, defined as all patients who received at least one dose of study treatment.

Pre-defined subgroup efficacy analyses were performed to compare treatments within the stratification factors, and between demographic and baseline characteristics. Table 4.3 presents a summary of the methodology for the AUGMENT trial.

Table 4.3: Summary of AUGMENT methodology

Trial Name	**AUGMENT**
Location	96 sites across 17 countries across North America, Europe, China and Brazil
Trial design	A multinational, randomised, double-blind, Phase III study Patients were randomised in a 1:1 ratio through an IVRS Randomisation was stratified by previous rituximab treatment (yes, no), time since last anti-lymphoma therapy (≤2, >2 years) and disease histology (FL, MZL)
Eligibility criteria for participants	**Inclusion criteria:** • Aged ≥18 years • Histologically confirmed MZL or Grade 1, 2, or 3a FL (CD20+ by flow cytometry or histochemistry) as assessed by investigator or local pathologist • Had to have been previously treated with at least one prior systemic chemotherapy, immunotherapy or R-chemo and had to have received at least two previous doses of rituximab: • Systemic therapy did not include local involved field radiotherapy for limited stage disease or *Helicobacter pylori* eradication • Prior investigational therapies were allowed provided the patient had received at least one prior systemic therapy • Had to have documented relapsed, refractory, or progressive disease (PD) after treatment with systemic therapy, and not be R-refractory • Rituximab-refractoriness was defined as did not respond (at least a PR) to rituximab or R-chemo therapy and/or time to disease progression <6 months after last rituximab dose • Rituximab-sensitive MZL or FL was defined as responded (at least a PR) to rituximab or R-chemo regimen therapy and time to disease progression ≥6 months after last rituximab dose

Trial Name	AUGMENT
	• Must have needed treatment for relapsed, progressed, or refractory disease as assessed by the investigator • Performance status ≤2 on the ECOG scale **Exclusion criteria:** • Life expectancy <6 months • Prior use of lenalidomide • Presence or history of central nervous system (CNS) involvement by lymphoma • Patients who were at a risk for a thromboembolic event and were not willing to take venous thromboembolism (VTE) prophylaxis
Settings and locations where the data were collected	An independent external DMC assessed ongoing safety throughout the study. The DMC conducted the planned interim futility analysis when an estimated 96 events per IRC review were reported. Response-related efficacy assessments were based on central review, including central radiology and clinical review by the IRC. Images received from investigators' sites were sent to the IRC, as well as relevant clinical information for haemato-oncology review.
Trial drugs	Lenalidomide 10 mg or 20 mg oral capsules[a] once daily on Days 1 to 21 of every 28-day Cycle up to 12 cycles combined with rituximab 375 mg/m^2 IV every week in Cycle 1 and on Day 1 of every 28-day Cycle from Cycles 2 through 5. Treatment continued until progression or unacceptable toxicity.
Permitted and disallowed medication	The following medications are prohibited during the study: • Systemic chronic corticosteroid at doses above 20 mg/day (prednisone/prednisolone or equivalent) during treatment phase. A seven-day washout period before Cycle 1 Day 1 study drug dosing was required for these patients • All investigational therapies (drug or otherwise) and anticancer therapies, other than lenalidomide or rituximab were prohibited during the entire Treatment Period of the study
Primary outcomes (including scoring methods and timings of assessments)	• PFS in relapsed/refractory indolent lymphoma patients, defined as the time from randomization to the first observation of disease progression, based on the modified 2007 IWGRC, or death due to any cause • Analysis was based on the IRC determination of disease progression

Other outcomes used in the economic model/specified in the scope	**Secondary endpoints**
	• To compare the safety of R^2 versus rituximab plus placebo
	• To compare the efficacy of R^2 versus rituximab plus placebo using other parameters of efficacy:
	  • DCRR, ORR, CR rate, DOR, and DOCR by the 2007 IWGRC without PET
	  • OS, EFS, and TTNLT
	Exploratory endpoints
	• To compare the effects of R^2 versus R-mono on:
	  • TTNCT and RTNLT
	  • CR/CRu rate in patients with FL based on the 1999 IWGRC
	  • PFS on next anti-lymphoma treatment (PFS2)
	  • HRQL as measured by the EORTC Quality of Life Questionnaire, Core 30 (QLQ-C30) and EuroQol Group's questionnaire 5 dimensions (EQ-5D-3L)

Pre-planned subgroups	Efficacy analyses were performed within a number of patient subgroups. These are described in Appendix M of the CS.

Trial Name	AUGMENT
Source: CS Table 4, pages 30-32. CR = complete response; CT = computerised tomography; DCRR = durable complete response rate; DMC = data monitoring committee; DOCR = duration of complete response; DOR = duration of response; ECOG = Eastern Cooperative Oncology Group; EFS = event-free survival; EORTC = European Organisation for Research and Treatment of Cancer; FL = follicular lymphoma; HRQL = health-related quality of life; IRC = Independent Review Committee; IVRS = interactive voice response system; IWGRC = International Working Group Response Criteria; MALT = mucosa-associated lymphoid tissue; MRI = magnetic resonance imaging; MZL = marginal zone lymphoma; ORR = overall response rate; PD = progressive disease; PR = partial response; R^2 = rituximab plus lenalidomide; R-chemo = rituximab-containing chemotherapy; R-mono = rituximab monotherapy; RTNLT = response rate to next anti-lymphoma treatment; TTNLT = time to next anti-lymphoma treatment; TTNCT = time to next chemotherapy treatment. Notes: [a] dose modification rules allowed for dosing down to 2.5 mg with Celgene supplying lenalidomide 2.5 mg, 5 mg, 10 mg, 15 mg, and 20 mg capsules.	

4.2.3 Baseline characteristics of the AUGMENT trial

Baseline characteristics for patients in the AUGMENT trial are presented in Table 4.4.

ERG comment: Of note, more patients in the R^2 arm than in the R-mono arm were female (58% vs. 46%), aged ≥65 years (46% vs. 41%) had Ann Arbor Stage III to IV disease (77% vs. 69%), FLIPI score ≥3 (39% vs. 30%), had an ECOG score of 1 or 2 (35% vs. 29%) and were refractory to the last prior regimen (17% vs. 14%). In addition, the company stated that for patients with MZL, baseline disease characteristics were imbalanced and favoured the R-mono arm (R^2 arm vs. R-mono arm): ECOG 0 (55% vs 72%); Ann Arbor Stage III to IV disease (77% vs. 56%); Ann Arbor Stage IV (65% vs. 41%); FLIPI score ≥3 (48% vs. 25%); B symptoms (13% vs. 3%); and high tumour burden per GELF criteria (65% vs. 56%). The ERG agrees with this and judged that the baseline characteristics for MZL patients may favour R-mono.

Table 4.4: Baseline demographic and disease characteristics, AUGMENT – ITT population

	FL		MZL		Total		Overall (n=358)
	R^2 (n=147)	**R-mono (n=148)**	**R^2 (n=31)**	**R-mono (n=32)**	**R^2 (n=178)**	**R-mono (n=180)**	
Male, n (%)	61 (41.5)	80 (54.1)	14 (45.2)	17 (53.1)	75 (42.1)	97 (53.9)	172 (48.0)
Median age, years (range)	62.0 (26.0-86.0)	61.0 (35.0-88.0)	68.0 (37.0-80.0)	66.0 (36.0-82.0)	64.0 (26.0-86.0)	62.0 (35.0-88.0)	62.5 (26.0-88.0)
Age distribution, n (%)							
<65	86 (58.5)	94 (63.5)	10 (32.3)	13 (40.6)	96 (53.9)	107 (59.4)	203 (56.7)
≥65	61 (41.5)	54 (36.5)	21 (67.7)	19 (59.4)	82 (46.1)	73 (40.6)	155 (43.3)
≥70	34 (23.1)	32 (21.6)	13 (41.9)	12 (37.5)	47 (26.4)	44 (24.4)	91 (25.4)
Race, white (%)	[redacted]	[redacted]	[redacted]	[redacted]	118 (66.3)	115 (63.9)	233 (65.1)
Histology (investigator review), n (%)							
FL	[redacted]	[redacted]	[redacted]	[redacted]	147 (82.6)	148 (82.2)	295 (82.4)
Grade 1	[redacted]	[redacted]	[redacted]	[redacted]	50 (28.1)	62 (34.4)	112 (31.3)
Grade 2	[redacted]	[redacted]	[redacted]	[redacted]	75 (42.1)	61 (33.9)	136 (38.0)
Grade 3a	[redacted]	[redacted]	[redacted]	[redacted]	22 (12.4)	25 (13.9)	47 (13.1)
MZL	N/A	N/A	31 (100.0)	32 (100.0)	31 (17.4)	32 (17.8)	63 (17.6)
MALT	N/A	N/A	14 (45.2)	16 (50.0)	14 (7.9)	16 (8.9)	30 (8.4)
Nodal	N/A	N/A	8 (25.8)	10 (31.3)	8 (4.5)	10 (5.6)	18 (5.0)
Splenic	N/A	N/A	9 (29.0)	6 (18.8)	9 (5.1)	6 (3.3)	15 (4.2)
Ann Arbor stage, n (%)							
I	13 (8.8)	13 (8.8)	2 (6.5)	5 (15.6)	15 (8.4)	18 (10.0)	33 (9.2)
II	21 (14.3)	29 (19.6)	5 (16.1)	9 (28.1)	26 (14.6)	38 (21.1)	64 (17.9)
III	69 (46.9)	60 (40.5)	4 (12.9)	5 (15.6)	73 (41.0)	65 (36.1)	138 (38.5)
IV	44 (29.9)	46 (31.1)	20 (64.5)	13 (40.6)	64 (36.0)	59 (32.8)	123 (34.4)

	FL		MZL		Total		Overall (n=358)
	R² (n=147)	R-mono (n=148)	R² (n=31)	R-mono (n=32)	R² (n=178)	R-mono (n=180)	
FLIPI category (derived), n (%)							
Low (0,1)	[redacted]	[redacted]	[redacted]	[redacted]	52 (29.2)	67 (37.2)	119 (33.2)
Intermediate (2)	[redacted]	[redacted]	[redacted]	[redacted]	55 (30.9)	58 (32.2)	113 (31.6)
High (≥3)	[redacted]	[redacted]	[redacted]	[redacted]	69 (38.8)	54 (30.0)	123 (34.4)
Baseline ECOG score, n (%)							
0	99 (67.3)	105 (70.9)	17 (54.8)	23 (71.9)	116 (65.2)	128 (71.1)	244 (68.2)
1	47 (32.0)	42 (28.4)	13 (41.9)	8 (25.0)	60 (33.7)	50 (27.8)	110 (30.7)
2	[redacted]	[redacted]	[redacted]	[redacted]	2 (1.1)	2 (1.1)	4 (1.1)
LDH elevated, n (%)							
Yes	39 (26.5)	43 (29.1)	6 (19.4)	6 (18.8)	45 (25.3)	49 (27.2)	94 (26.3)
No	107 (72.8)	105 (70.9)	25 (80.6)	26 (81.3)	132 (74.2)	131 (72.8)	263 (73.5)
High tumour burden (GELF criteria)							
Yes	77 (52.4)	68 (45.9)	20 (64.5)	18 (56.3)	97 (54.5)	86 (47.8)	183 (51.1)
No	70 (47.6)	80 (54.1)	11 (35.5)	14 (43.8)	81 (45.5)	94 (52.2)	175 (48.9)
Prior anti-lymphoma regimens							
1	[redacted]	[redacted]	[redacted]	[redacted]	102 (57.3)	97 (53.9)	199 (55.6)
>1	[redacted]	[redacted]	[redacted]	[redacted]	76 (42.7)	83 (46.1)	159 (44.4)
Refractory to last prior regimen							
Yes	26 (17.7)	25 (16.9)	4 (12.9)	1 (3.1)	30 (16.9)	26 (14.4)	56 (15.6)
No	121 (82.3)	123 (83.1)	27 (87.1)	31 (96.9)	148 (83.1)	154 (85.6)	302 (84.4)
POD24[a], n (%)							
Yes	[redacted]	[redacted]	[redacted]	[redacted]	56 (31.5)	61 (33.9)	117 (32.7)
No	[redacted]	[redacted]	[redacted]	[redacted]	122 (68.5)	118 (65.6)	240 (67.0)

	FL		MZL		Total		Overall (n=358)
	R^2 (n=147)	R-mono (n=148)	R^2 (n=31)	R-mono (n=32)	R^2 (n=178)	R-mono (n=180)	

Source: CS, Table 5, pages 34-35.

ECOG = Eastern Cooperative Oncology Group; FL = follicular lymphoma; FLIPI = follicular lymphoma international prognostic index; GELF = Groupe d'Etude des Lymphomes Folliculaires; LDH = lactate dehydrogenase; MZL = marginal zone lymphoma; MALT = mucosa associated lymphatic tissue; R^2 = lenalidomide plus rituximab; R-mono = rituximab plus placebo.

Notes: [a]) POD24 is defined as relapse within two years of initial chemoimmunotherapy.

4.2.4 Statistical analyses of the AUGMENT trial

The primary outcome of AUGMENT was PFS. The primary analysis was performed in the ITT population using outcomes assessed by the IRC using a modified version of the 2007 IWGRC. Analyses were performed using both FDA and European Medicines Agency (EMA) censoring rules for PFS but only the EMA censoring rule analyses for the ITT population were presented in the main body of the CS. Safety assessments for the study were conducted on the safety population.

Table 4.5 presents the hypothesis and associated statistical analysis methods adopted in the AUGMENT trial. PFS was defined as the time from the date of randomisation to the first observation of documents disease progression or death from any cause, whichever occurred first. The analysis compared Kaplan-Meier survival curves using a log-rank test (one sided $p < 0.025$) and a Cox proportional hazards model. OS was also analysed using Kaplan-Meier estimates of OS.

Overall response rate (ORR) was defined as the proportion of patients with best response of at least PR without administration of new anti-lymphoma therapy. Complete response (CR) was the proportion of patients with a best response of CR during the study without administration of new anti-lymphoma therapy. ORR and CR were compared between treatment groups using a stratified Cochran-Mantel-Haenszel (CMH) test stratified by the randomisation stratification factors.

Planned subgroup analyses included the randomisation stratification factors previous rituximab treatment (yes, no), time since last anti-lymphoma therapy (≤ 2, >2 years), and histology (FL, MZL) and also age (<65, ≥65 years); gender (male, female); race (White; Other races); region (US, EU, Asia-Pacific region and Brazil); FLIPI (<3, ≥3) for FL patients only; number of prior anti-lymphoma regimens (1, >1); Ann Arbor stage at enrolment (I to II, III to IV); prior rituximab-containing chemotherapy regimen (yes, no); refractory to last prior regimen (defined as <PR or PD within six months from last systemic regimen) (yes, no); High tumour burden per Groupe d'Etude des Lymphomes Folliculaires (GELF) criteria (yes, no); chemo-resistant (<PR or PD within six months from last chemotherapy) (yes, no) or ECOG performance status ≥2 [yes; no])

ERG comment: The statistical analysis of the trial used appropriate methods and the ERG does not have any concerns.

Table 4.5: Summary of statistical analyses

Study	Hypothesis objective	Statistical analysis	Sample size, power calculation	Data management, patient withdrawals
AUGMENT	The primary objective of the study was to compare the efficacy of R^2 to R-mono. Efficacy determination was based on PFS as the primary endpoint. The AUGMENT study was considered positive if the R^2 group was significantly superior to the rituximab group for the primary endpoint.	The analysis of the primary endpoint was planned when approximately 193 IRC-assessed PFS events were reached. The cut-off date for database lock was prespecified before database lock. KM estimates of PFS were provided, and the KM product limit method was used to estimate the survivorship function for PFS. Event rates at specific time points were estimated from KM curves. Medians together with two-sided 95% CIs were provided. The resulting PFS estimates were presented graphically.	Based on the rate of accrual anticipated in this study and 5% annual dropout rate, it was estimated that approximately 350 patients would be randomised in a 1:1 ratio to the two treatment arms and that PFS would be reached at 43 months. The basis for the power and sample size determination was a test of the equality of the overall time-to-event (i.e. PFS) curves between experimental and control treatment groups using a stratified log-rank test.	**EMA censoring rules** Event: • Death before first PD assessment while on study • Death between adequate assessment visits • All progressions and deaths, regardless of whether they occurred after next anti-lymphoma therapy or after ≥2 missed scheduled assessments Censored: • Patients with no baseline assessment were censored at randomisation • Patients who did not progress or die and those that discontinued for any reason other than death or progression will be censored on the date of their last adequate assessment with evidence of no progression • Patients who died or progressed after more than one missed visit will be censored at the date of their last adequate assessment that revealed no progression

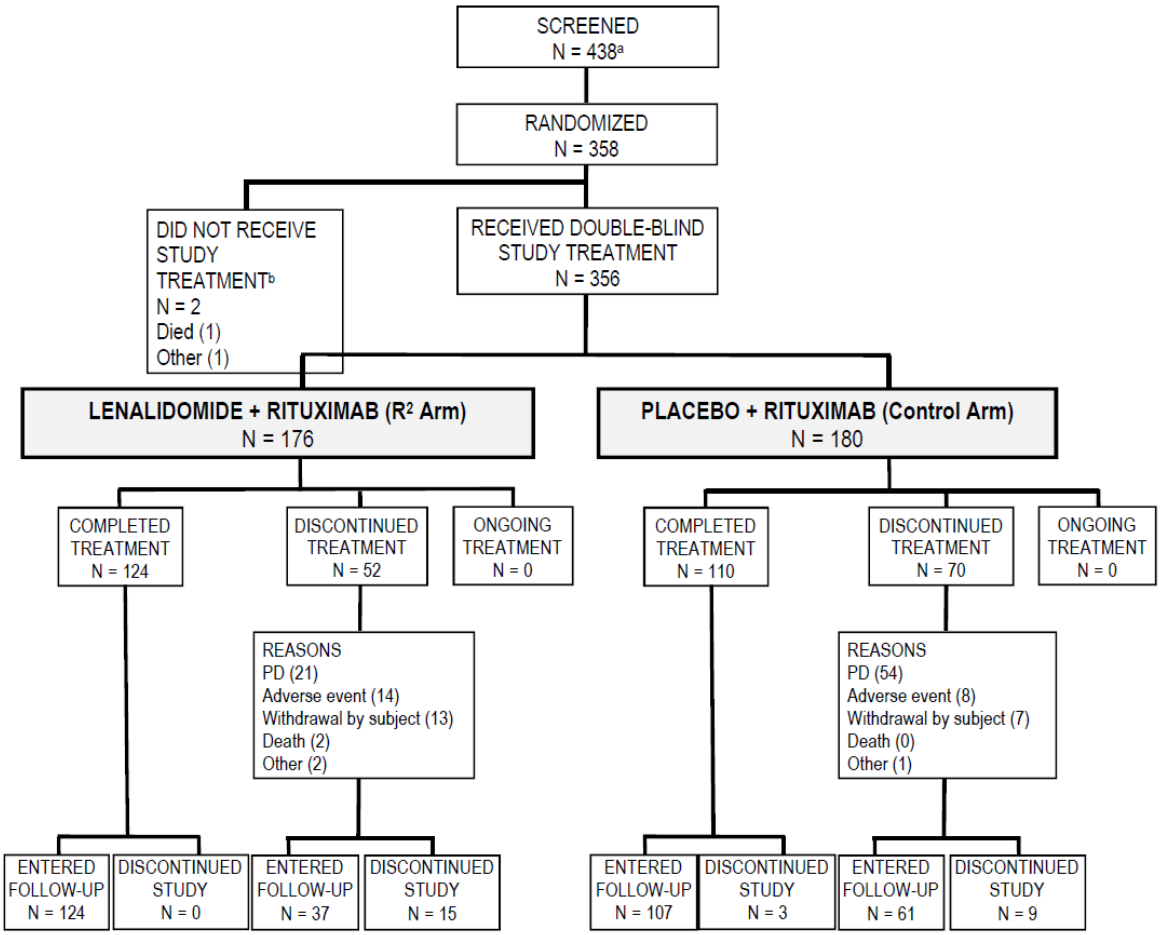
Source: CS, Table 8, pages 46-48.
CI = confidence interval; EMA = European Medicines Agency; IEE = induction efficacy population; IRC = Independent Review Committee; ITT = intention-to-treat; KM = Kaplan–Meier; ORR = overall response rate; PD = progressive disease; PFS = progression-free survival; R^2 = lenalidomide plus rituximab; R-mono = rituximab plus placebo.

4.2.5 Results of the AUGMENT trial

The data presented in the CS are based on the 22 June 2018 data cut-off for the primary analysis. Efficacy analyses were conducted in the ITT population and based on data from IRC review, using the modified 2007 IWGRC. EMA censoring rules were applied to the analyses.

At the time of the data cut-off (22 June 2018) more patients in the R^2 arm had completed treatment compared with the R-mono arm. In the R^2 arm, 124 patients (70.5%) had completed treatment, 52 patients (29.5%) had discontinued treatment, and no patients were ongoing with treatment. In the R-mono arm, 110 patients (61.1%) had completed treatment, 70 patients (38.9%) had discontinued treatment, and no patients were ongoing with study treatment (see Figure 4.1).

Figure 4.1: CONSORT diagram of patient flow during the AUGMENT trial



Source: CS, Appendix D4, pages 59-60.

FL = follicular lymphoma; MZL = marginal zone lymphoma; PD = progressive disease.

Notes: [a]) in total, 438 patients were screened for study participation, of which 18 patients (4.1%) were screened twice. Of the total 456 screens, 98 were screen failures primarily due to failure of inclusion/exclusion criteria (96.9%). Screen failures either did not meet inclusion criteria (n=70) and/or met at least one exclusion criterion (n=28); [b]) two patients randomised to the R^2 arm did not receive study medication: one patient with MZL died due to septic shock after randomisation but prior to receiving the first dose of study treatment and one patient with FL discontinued due to Grade 2 dyspnoea on Cycle 1 Day 1, prior to administration of the first dose of study drug.

The overall median follow-up time for surviving patients in the ITT Population was 28.30 months (range: 0.1 to 51.3 months); this was comparable between FL and MZL patients.

Table 4.6 presents a summary of the main results from the AUGMENT trial. Results for FL and MZL separately are reported in Appendix 1 of this report.

Table 4.6: Summary of results from the AUGMENT trial: ITT population

Endpoint	**Overall**	
	R^2 (n=178)	**R-mono (n=180)**
Median OS, months (95% CI)[a]	NE (NE, NE)	NE (NE, NE)
Hazard ratio (95% CI)	0.61 (0.33, 1.13)[b]	
Median PFS, months (95% CI)[a]	[redacted]	[redacted]
Hazard ratio (95% CI)	[redacted][b]	
Best response, n (%)		
ORR (CR+PR)	138 (77.5)	96 (53.3)
95% CI[d]	70.7, 83.4	45.8, 60.8
p-value	<0.0001[e]	
CR rate	60 (33.7)	33 (18.3)
95% CI[d]	26.8, 41.2	13.0, 24.8
p-value	0.001[e]	
PR	78 (43.8)	63 (35.0)
SD	20 (11.2)	55 (30.6)
PD/ death	7 (3.9)	23 (12.8)
No evidence of disease	3 (1.7)	4 (2.2)
Unknown/ND/Missing	10 (5.6)	2 (1.1)
Median TTNLT, months (95% CI)[a]	NE (NE, NE)	32.2 (23.2, NE)
TTNLT rate at 2 years, % (95% CI)	[redacted]	[redacted]
Hazard ratio (95% CI)	0.54 (0.38, 0.78)[b]	
p-value	0.0007[g]	
Median EFS, months (95% CI)[a]	27.6 (22.1, NE)	13.9 (11.4, 16.7)
Hazard ratio (95% CI)	0.51 (0.38 to 0.67)[b]	
p-value	<0.0001[g]	
Median TTNCT, months (95% CI)[a]	NE (NE, NE)	NE (NE, NE)
TTNCT rate at 2 years, % (95% CI)	[redacted]	[redacted]
Hazard ratio (95% CI)	0.50 (0.32, 0.78)[b]	
p-value	0.0017[g]	
RTNLT		
ORR, n (% [95% CI][d])	28 (57.1 [42.2, 71.2])	29 (36.3 [25.8, 47.8])
p-value	0.0282[f]	
CR, n (% [95% CI][d])	15 (30.6 [18.3, 45.4])	13 (16.3 [8.9, 26.2])
p-value	0.0775[f]	
DCRR, n/N (%)	[redacted]	[redacted]

95% CI[d]	[redacted]	[redacted]
p-value	[redacted][e]	
N, Median DOR, months (95% CI)[a]	[redacted]	[redacted]
Hazard ratio (95% CI)[c]	0.53 (0.36 to 0.79)	
p-value[e]	0.0015	
N, Median DOCR, months (95% CI)[a]	60, NE (25.3, NE)	33, NE (13.8, NE)
Hazard ratio (95% CI)[h]	[redacted]	
p-value	[redacted]	

Source: CS, Table 10 and 11, pages 52 and 55-56.
CI = confidence interval; CR = complete response; DCRR = durable complete response rate, DOCR = duration of complete response; DOR = duration of response; EFS = event-free survival; FL = follicular lymphoma; IRC = Independent Review Committee; ITT = intent-to-treat; MZL = marginal zone lymphoma; ND = not done; NE = not estimable; ORR = overall response rate; OS = overall survival; PD = progressive disease; PR = partial response; R^2 = lenalidomide plus rituximab; R-mono = rituximab plus placebo; RTNLT = response rate to next anti-lymphoma treatment; SD = stable disease; TTNLT = time to next anti-lymphoma treatment; TTNCT = time to next anti-lymphoma chemotherapy treatment.
Notes: [a]) median estimate is from Kaplan–Meier analysis; [b]) from Cox proportional hazard model adjusting for the three stratification factors: previous rituximab treatment (yes; no), time since last anti-lymphoma therapy (≤2; >2 year), and disease histology (FL; MZL). [c]) from Cox proportional hazard model; [d]) exact confidence interval for binomial distribution; [e]) from CMH test adjusting for the three stratification factors; [f]) from Fisher-Exact test; [g]) from log-rank test adjusting for the three stratification factors; [h]) from log-rank test; [i]) exact confidence interval for binomial distribution.

Overall, R^2 showed favourable results when compared to R-mono for PFS with a greater median PFS ([redacted] months; HR [redacted] (95% CI: [redacted]). However, there was no evidence of a difference in overall survival (OS) with a hazard ratio (HR) of 0.61 (95% CI: 0.33 to 1.13) for patients treated with R^2 compared to R-mono. At the time of the analysis the OS data was immature with 16 deaths on R^2 and 26 deaths on R-mono at the time of the analysis. Overall response rate (ORR) was significantly greater for R^2 compared with R-mono (78% vs. 53%; p<0.0001). The complete response (CR) rate was also greater for the R^2 arm compared with R-mono (34% vs. 18%; p=0.001).

ERG comment: As can be seen from Tables A1.1 and A1.2 (see Appendix 1 of this report), results for R^2 versus R-mono in MZL patients are generally less favourable for R^2 than in FL patients. However, it is important to note that PFS outcomes in the MZL subgroup are difficult to interpret because of the small sample size (31 patients in the R^2 arm and 32 patients in the R-mono arm) and imbalance in baseline prognostic factors (as discussed in Section B.2.3 of the CS and section 4.2.3 of this report).

Health-related quality of life (HRQoL)

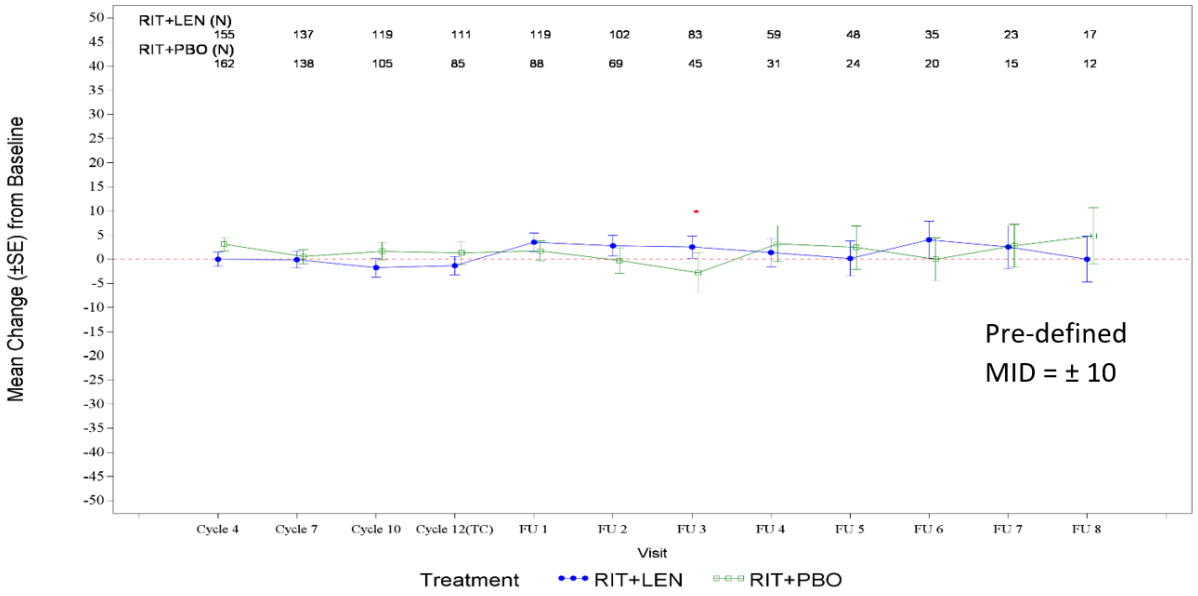
HRQoL was assessed using the EORTC QLQ-C30 and EuroQol Five Dimension Three Level (EQ-5D-3L) questionnaire. The global health status/quality of life (GHS/QoL) domain of the QLQ-C30 was chosen as the primary patient reported outcome of interest.

Primary HRQoL analyses were performed on the HRQoL-evaluable population, defined as patients in the ITT population who had a GHS/QoL domain score at baseline and at least one post-baseline assessment. The ITT population was also analysed, but only to assess the HRQoL compliance rates. The HRQoL-evaluable population comprised of 338 patients (94% of the ITT population), including 165 patients receiving R^2 and 173 patients receiving R-mono.

A minimal important difference (MID) of a ≥10-point change from baseline at the individual patient level was used to define the proportion of patients reporting a meaningful difference in QOL for any given domain of the EORTC QLQ-C30.

Based on the results from the cross-sectional analysis (within- and between-group difference in mean change from baseline score at each post-baseline assessment visit), no clinically meaningful change from baseline in the GHS/QoL domain of the QLQ-C30 was observed across any of the post-baseline assessment visits, regardless of treatment group (See Figure 4.2). Between-group differences in mean changes were small and not clinically meaningful across all assessment visits and did not differ between FL and MZL patients. Furthermore, change from baseline scores over time, based on the cross-sectional assessment, did not differ meaningfully by response status, occurrence of Grade 3/4 AEs, and occurrence of any neutropenia. The longitudinal assessment also indicated no statistically significant or clinically meaningful difference in LS mean changes from baseline between treatment groups across all timepoints; and no change exceeded the MID threshold.

Figure 4.2: Cross-sectional assessment of global health status/quality of life changes from baseline



Source: CS, Appendix P, Figure 22, page 237.
FU = follow-up; MID = minimally important difference; Len = lenalidomide; PBO = placebo; Rit = rituximab; SE = standard error; TC = treatment completion.

4.2.6 Adverse events

Adverse event data from the AUGMENT trial were taken from the 22 June 2018 database cut-off; safety analyses were conducted in the safety population.

Overall, the median lenalidomide/placebo treatment duration was [redacted] months for the R^2 arm and [redacted] months for the R-mono arm. The median rituximab treatment duration was also similar between the R^2 and R-mono arms ([redacted] vs. [redacted], respectively).

A summary of the treatment-emergent adverse event (TEAEs) during AUGMENT for the total population (FL and MZL) is presented in Table 4.7. TEAEs were reported in 174 patients (99%) in the R^2 arm and 173 patients (96%) in the R-mono arm. More patients in the R^2 arm (69%) experienced a Grade 3 or 4 TEAE compared with those in the R-mono arm (32%), and two patients in each treatment arm reported a Grade 5 TEAE. Additionally, a greater proportion of patients reported serious adverse events in the R^2 arm (26%) compared with those in the R-mono arm (14%). Separate tables for FL and MZL patients are presented in Appendix 1 of this report.

Table 4.7: Summary of treatment-emergent adverse events in AUGMENT: Safety population

	Total population (FL + MZL)	
	R^2 (n=176)	**R-mono (n=180)**
Number of patients (%)		
Any TEAE	174 (98.9)	173 (96.1)
Len related	159 (90.3)	118 (65.6)
R related	132 (75.0)	105 (58.3)
Grade 3–4 TEAE	121 (68.8)	58 (32.2)
Len related	101 (57.4)	38 (21.1)
R related	57 (32.4)	19 (10.6)
Grade 5 TEAE	2 (1.1)	2 (1.1)
Any SAE	45 (25.6)	25 (13.9)
Len related	23 (13.1)	8 (4.4)
R related	13 (7.4)	3 (1.7)
Any TEAE leading to dose reduction of Len/Pbo	46 (26.1)	6 (3.3)
Any TEAE leading to dose interruption of Len/Pbo	112 (63.6)	47 (26.1)
Any TEAE leading to dose interruption of R	60 (34.1)	37 (20.6)
Any TEAE leading to discontinuation of Len/Pbo	15 (8.5)	9 (5.0)
Any TEAE leading to discontinuation of R	6 (3.4)	2 (1.1)

Source: CS, Table 21, page 94 and Clarification Letter, Table 6, page 21.
Len = lenalidomide; Pbo = placebo; R = rituximab; R2 = lenalidomide + rituximab; R mono = placebo, rituximab + placebo; SAE = serious adverse event; TEAE = treatment-emergent adverse event.

In the safety population, TEAEs that occurred more frequently (≥10% difference) in the R^2 arm than the R-mono arm included the following: neutropenia (58% vs. 22%), diarrhoea (31% vs. 23%), constipation (26% vs. 14%), cough (23% vs. 17%), upper respiratory tract infection (18% vs. 13%) and leukopenia (20% vs. 9%) (see Table 4.8).

The difference in the number of Grade 3 or 4 TEAEs between treatment arms (shown in Table 4.7) was largely driven by Grade 3 or 4 events of neutropenia and leukopenia. Neutropenia occurred in 88 patients (50%) in the R^2 arm compared with 23 patients (13%) in the R-mono arm, and leukopenia occurred in 12 patients (7%) in the R^2 arm compared with three patients (2%) in the R-mono arm.

The most common TEAEs, occurring in more than 10% of patients, are presented in Table 4.8 below. Separate adverse events tables for FL and MZL patients are presented in Appendix 1 of this report.

Table 4.8: Most common treatment-emergent adverse events reported in ≥10% of patients in either treatment arm in AUGMENT: Safety population (FL and MZL)

	Total population (FL + MZL)	
	R^2 (n=176)	**R-mono (n=180)**
Number of patients (%)		
Blood and lymphatic system disorders	118 (67.0)	58 (32.2)
Neutropenia	102 (58.0)	40 (22.2)
Leukopenia	36 (20.5)	17 (9.4)
Anaemia	28 (15.9)	8 (4.4)
Thrombocytopenia	26 (14.8)	8 (4.4)
Gastrointestinal disorders	115 (65.3)	88 (48.9)
Diarrhoea	55 (31.3)	41 (22.8)
Constipation	46 (26.1)	25 (13.9)
Abdominal pain	22 (12.5)	16 (8.9)
Nausea	20 (11.4)	23 (12.8)
Infections and infestations	110 (62.5)	88 (48.9)
URTI	32 (18.2)	23 (12.8)
Nasopharyngitis	13 (7.4)	18 (10.0)
General disorders and administration site conditions	98 (55.7)	89 (49.4)
Fatigue	38 (21.6)	33 (18.3)
Pyrexia	37 (21.0)	27 (15.0)
Asthenia	24 (13.6)	19 (10.6)
Oedema peripheral	23 (13.1)	16 (8.9)
Skin and subcutaneous tissue disorders	89 (50.6)	43 (23.9)
Pruritus	21 (11.9)	7 (3.9)
Rash	19 (10.8)	7 (3.9)
Musculoskeletal and connective tissue disorders	73 (41.5)	58 (32.2)
Muscle spasms	23 (13.1)	9 (5.0)
Back pain	14 (8.0)	18 (10.0)
Respiratory, thoracic and mediastinal disorders	73 (41.5)	65 (36.1)
Cough	40 (22.7)	31 (17.2)
Dyspnoea	19 (10.8)	8 (4.4)
Investigations	60 (34.1)	50 (27.8)
Alanine aminotransferase increased	18 (10.2)	15 (8.3)
Metabolism and nutrition disorders	58 (33.0)	40 (22.2)
Decreased appetite	23 (13.1)	11 (6.1)
Nervous system disorders	58 (33.0)	39 (21.7)
Headache	26 (14.8)	17 (9.4)
Injury, poisoning and procedural complications	42 (23.9)	40 (22.2)
Infusion related reaction	26 (14.8)	24 (13.3)

	Total population (FL + MZL)	
	R^2 (n=176)	**R-mono (n=180)**
Eye disorders	28 (15.9)	14 (7.8)
Neoplasms benign, malignant and unspecified (including cysts and polyps)	26 (14.8)	9 (5.0)
Tumour flare	19 (10.8)	1 (0.6)
Psychiatric disorders	24 (13.6)	20 (11.1)
Cardiac disorders	21 (11.9)	17 (9.4)
Vascular disorders	21 (11.9)	22 (12.2)

Source: CS, Appendix F, Table 31, pages 63-64 and Clarification Letter, Table 7, pages 22-23.
R^2 = lenalidomide + rituximab; R-placebo = rituximab + placebo; URTI = upper respiratory tract infection.

ERG comment: As shown in Table 4.7, R^2 was associated with more grade 3-4 TEAEs and SAEs when compared to R-mono, especially lenalidomide/placebo related adverse events; but rituximab-related grade 3-4 TEAEs and SAEs were also more frequent in the R^2 arm than in the R-mono arm. R^2 was also associated with more TEAEs leading to dose reductions, dose interruptions and discontinuations of lenalidomide/placebo or rituximab when compared to R-mono. Adverse events are generally the same for FL and MZL patients; however, AEs for MZL patients are based on small numbers.

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

The systematic literature review (SLR) performed by the company identified 45 studies (13 RCTs and 32 non-RCTs). According to the company, 39 studies were considered not relevant for the submission because they did not investigate comparators of interest (lenalidomide (1x), obinutuzumab plus lenalidomide (1x), idelalisib (4x), copanlisib (2x), ibrutinib (3x), rituximab plus bendamustine (6x), other bendamustine-containing regimens (1x), rituximab monotherapy (15x), bendamustine monotherapy (5x), and tazemetostat (1x)). Therefore, the company included a total of six relevant studies.

ERG comment: As explained in Section 3.3 of this report, rituximab monotherapy is a relevant comparator for this appraisal according to the NICE scope. Therefore, the 15 studies investigating rituximab monotherapy should have been included. However, as there is a trial with a head-to-head comparison of R^2 with rituximab monotherapy, the 15 rituximab monotherapy studies can probably be ignored.

Of the six relevant studies identified by the company, there were five relevant RCTs (AUGMENT (R^2),[35] MAGNIFY (R^2),[36] ALLIANCE (R^2),[37] Van Oers (R-CHOP)[39] and GADOLIN (O-Benda)[40]) and one relevant non-RCT (Tuscano 2014[38] (R^2)). The SLR found no studies for the relevant comparator R-CVP.

ERG comment: The four R^2 studies were discussed in Section 4.1.2 of this report. This ERG report will focus on the AUGMENT trial, because this provides a head-to-head comparison of the intervention of interest (lenalidomide in combination with rituximab, R^2) versus a relevant comparator according to the NICE scope (rituximab monotherapy). The study by Van Oers et al. (2006), was relevant for the indirect comparison using published data and will be discussed in Section 4.4.1 of this report. The GADOLIN study was used by the company for an indirect comparison of R^2 with O-Benda. However, as explained in Sections 3.3 and 4.4 of this report, O-Benda is not considered by NICE to be a relevant comparator for this appraisal; therefore, this study will be ignored.

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

The company performed two types of indirect comparisons. First, the company performed an indirect comparison with data from published evidence. This included comparisons of R^2 with:

- R-CHOP for non-rituximab refractory patients, based on comparator data from a study by Van Oers et al. (2006)[39] comparing R-CHOP with CHOP (only the R-CHOP arm was used in the analyses).
- Established clinical management without lenalidomide, i.e. O-Benda for rituximab refractory patients, based on comparator data from a study by Sehn et al. (2016)[40] comparing O-Benda with bendamustine monotherapy (only the O-Benda arm was used in the analyses).

ERG comment: As explained in Section 3.3 of this report, NICE does not consider O-Benda a relevant comparator for disease that is refractory to rituximab. Therefore, this comparison will be ignored.

In the response from NICE to comments on the draft scope, NICE stated that "obinutuzumab in combination with bendamustine is only used as part of the Cancer Drugs Fund therefore it is not considered a relevant comparator for disease that is refractory to rituximab." (see NICE Response to comments on draft scope (page 4).[28] When NICE recommends a drug for use within the Cancer Drugs Fund (CDF), NICE considers that there is plausible potential for the drug to satisfy the criteria for routine commissioning, but there is significant remaining clinical uncertainty which needs more investigation, through data collection in the NHS or clinical studies.[27] This means that the cost effectiveness of drugs recommended for use within the CDF has not yet been established. Therefore, any comparisons of effectiveness or cost effectiveness with CDF-drugs are equally uncertain.

Second, the company performed an indirect comparison with data from the Haematological Malignancy Research Network (HMRN). This included a comparison of R^2 with:

- Pooled data for R-CHOP/R-CVP for non-rituximab refractory patients.

There was no data for R-refractory patients receiving O-Benda in the HMRN database (due to this regimen only being recently available) and so this data source was not used for this population.

ERG comment: The company stated that 'Due to small patient numbers for non-R-refractory patients receiving R-CHOP and R-CVP in the HMRN database, clinical expectation that R-CHOP and R-CVP would have similar efficacy in a relapsed/refractory setting and empirical data demonstrating this to be the case, efficacy analyses compared R^2 to the pooled R-CHOP/R-CVP' (CS, page 84).[1] The ERG was not convinced by this statement, and asked the company to provide further clarification (Clarification Letter, Question A14).[26] The company responded that 'Data for R-CHOP and R-CVP have been pooled given clinical feedback that it is not unreasonable to assume similar efficacy between R-CHOP and R-CVP in the relapsed/refractory setting, and HMRN clinical data supporting this'.[26] However, looking at the Advisory Board document provided by the company,[19] no such statement is included; therefore, it is not clear how this clinical feedback was obtained. In addition, clinicians did advise that '[redacted]', while '[redacted]'.[19] This suggests that R-CHOP and R-CVP are generally considered for different types of patients, making a comparison of the effectiveness of the two drugs problematic. Clinicians also advised that '[redacted]'.[19]

The company also provided data from the HMRN database, to show that R-CHOP and R-CVP have similar effectiveness. However, these data are based on small numbers of patients (63 in total; [redacted] for

R-CHOP and [redacted] for R-CVP). Analyses of OS and PFS using Cox proportional hazards models showed no significant difference between treatments after adjusting for other covariates (age, prior lines of therapy, early relapse, stage, nodal sites and prior rituximab). However, an analysis of a small sample which shows no statistically significant differences between the two treatments does not mean that one can infer they are equivalent and can be combined for further indirect comparisons. Analysis of a larger dataset with sufficient statistical power could lead to a different conclusion. The one covariate that was consistently related to outcome was age, which suggests that R-CVP will be more often considered for elderly patients and R-CHOP will be more often considered for younger patients; which means that the drugs are generally considered for different populations, making a comparison problematic. In conclusion, the ERG does not think the company has presented convincing evidence suggesting that R-CHOP and R-CVP have similar clinical effectiveness.

In the next two sections a critique of the two types of MAIC will be presented: using published evidence and using HMRN data.

4.4.1 MAIC comparing R^2 with R-CHOP based on published evidence.

Table 4.9 shows a list of potential effect modifiers/prognostic variables (EM/PVs) that would ideally be adjusted for in a matching-adjusted indirect comparison (MAIC), as identified and validated by external clinical experts consulted by the company.[19]

Table 4.9: Potential EM/PVs that would ideally be adjusted for in a MAIC

Characteristic	highest priority	Included in MAIC	Comments
Previous exposure to rituximab	Yes		Not included in MAIC Was 0% in Van Oers
FLIPI components: - Age (median if mean no reported) - Ann Arbor Stage (III-IV) - Nodal sites (>4) - High LDH	Yes	Yes Yes	 No data reported in Van Oers Not included in MAIC
Refractory to last therapy	Yes	Yes	
Prior lines of therapy 1 vs. 2 vs. >2	Yes	Yes (2 and 3+)	One prior line of therapy was not included
FLIPI risk group (low vs. intermediate vs. high)	Yes	Yes (medium and high)	Low FLIPI risk was not included
FLIPI2+ components: - Serum beta-2 microglobulin high - Bone marrow involvement - Diameter of largest node >6 cm - Haemoglobin <12 dL/L			 No data reported in Van Oers Not included in MAIC No data reported in Van Oers No data reported in Van Oers
Time from last treatment			No data reported in Van Oers
POD24			No data reported in Van Oers
ECOG performance status (0–1 vs. 2+)			No data reported in Van Oers
Presence of B-symptoms			Not included in MAIC
Source: CS, Section B.2.9, pages 70-71.			

EM = effect modifiers; ESS = effective sample size; FL = follicular lymphoma; FLIPI = Follicular Lymphoma International Prognostic Index; MAIC = matching-adjusted indirect comparison; MZL = marginal zone lymphoma; PV = prognostic variables; R = rituximab; R^2 = lenalidomide plus rituximab; R-CHOP = rituximab plus cyclophosphamide, doxorubicin, vincristine, prednisolone.
Notes: Adjusted N is the sum of the absolute weights. The patient characteristics presented are the potential EM/PVs that were included in the matching. The following potential EM/PVs had data for all included studies but were dropped from the matching to maintain a sufficiently large effective sample size for subsequent analysis: % previous rituximab exposure. The ESS and adjusted N including these variables were 4.2 and 0.1.

The company stated that 'if the adjustment resulted in an expected sample size and/or adjusted number of patients that was too small for analysis, then the list of variables used for adjustment was reduced before analysis. This was done to maintain the maximum number of the most clinically important variables in the adjustment. Several combinations of variables were explored. However, note that excluding known imbalanced covariates from matching may result in populations with differing levels of effect modifiers/prognostic variables on each treatment, which can bias the analysis results' (CS, page 71).[1]

Clinical advisors consulted by the company,[19] agreed that the most significant factor to be considered for MAIC of AUGMENT and MAGNIFY compared with comparator studies was prior rituximab exposure. Other important factors noted by advisors were FLIPI score, age, refractoriness to last therapy, duration of prior response and number of prior therapies. 'If inclusion of one or more of these factors in the MAIC is not possible, particularly with respect to prior rituximab experience, or where their application sufficiently reduces effective sample size, the credibility of comparison of the rituximab-non-refractory patient data from AUGMENT/MAGNIFY with published data for R-CHOP and R-bendamustine would be limited.'[19]

ERG comment: Clinical advisors agreed that the most significant factor to be considered for MAIC was prior rituximab exposure, yet this could not be included in the MAIC as none of the patients in Van Oers had prior rituximab. Therefore, the credibility of comparison of the rituximab non-refractory patient data from AUGMENT with published data for R-CHOP is limited, according to the clinical advisors consulted by the company. Previous rituximab use was one of the major exclusion criteria in the study by Van Oers et al. (2006).[39] That means that all patients in Van Oers et al. are 100% rituximab-naïve and that the study is not reflective of UK practice, as acknowledged by the company (CS, page 101).[1] Several covariates were not included in the MAIC because data were not reported in the study by Van Oers et al. (2006).[39] Although this is through no fault of the company, it affects the reliability of the MAIC results as all possible covariates present in both studies should be adjusted for.

Standard methods for MAIC were used as recommended in NICE DSU TSD report 18.[41] Individual patient data (IPD) from AUGMENT and summary data from the Van Oers et al. (2006) study[39] (for rituximab-naïve FL patients only) were used for the comparisons in the non-R-refractory population. The IPD from AUGMENT was matched to the R-CHOP data to ensure similar baseline characteristics using recommended weighting methods. The matching used the maximum set of covariates (based on what was available in both studies but excluding previous rituximab exposure).

For the analysis of OS and PFS using the matched data, pseudo-IPD data were generated from the published KM curves using the Guyot method for digitising curves.[42] This data was compared to the IPD survival data for R^2 using a number of statistical methods: KM curves, a Cox proportional hazards model; and different parametric survival models (exponential, Weibull, Gompertz, log-normal, log-logistic and generalised gamma). The proportional hazards assumption and underlying assumptions of the parametric models were assessed.

The results of the matching for the EMs/PVs included in the matching are provided in Table 4.10 for all covariates included.

Table 4.10: Patient characteristics, observed and match-adjusted for the non-R-refractory population (FL and MZL), comparing R^2 (AUGMENT) and R-CHOP (Van Oers 2006)

Characteristic	AUGMENT (R^2) (n=178)	Van Oers (R-CHOP) (n=234)	Adjusted R^2 (n=78.8)
Patient characteristics			
% refractory	16.9	16.0	16.0
% Ann Arbor stage III-IV	77.0	100.0	100.0
% FLIPI medium	30.9	33.0	33.0
% FLIPI high	38.8	37.0	37.0
% 2 prior lines of therapy	17.4	22.0	22.0
% 3+ prior lines of therapy	25.3	0.0	0.0
Age	62.3	54.0	54.0
Outcomes			
OS	Not estimable	NR	NR
HR (95% CI)		[redacted]	
PFS (N, median (95% CI))	178, 39.4 months (NR)	234, 33.1 months (NR)	78.8, 30.4 months (NR)
HR (95% CI)		[redacted]	

Source: CS, Appendix D2, Table 15, page 36.
EM = effect modifiers; ESS = effective sample size; FL = follicular lymphoma; FLIPI = Follicular Lymphoma International Prognostic Index; MZL = marginal zone lymphoma; PV = prognostic variables; R = rituximab; R^2 = lenalidomide plus rituximab; R-CHOP = rituximab plus cyclophosphamide, doxorubicin, vincristine, prednisolone.
Notes: Adjusted N is the sum of the absolute weights. The patient characteristics presented are the potential EM/PVs that were included in the matching. The following potential EM/PVs had data for all included studies but were dropped from the matching to maintain a sufficiently large effective sample size for subsequent analysis: % previous rituximab exposure. The ESS and adjusted N including these variables were 4.2 and 0.1.

ERG comment: The comparison of R^2 versus R-CHOP was adjusted for the variables listed in Table 4.10, i.e. percentage of patients that were refractory, Ann Arbor score and FLIPI score, prior lines of therapy and age. Most of these variables were already reasonably balanced between group, except the percentage of patients with three or more prior lines of therapy, which was 0% in the study by Van Oers and 25% in AUGMENT. The MAIC results for the comparison of R^2 versus R-CHOP are only applicable to the population of Van Oers et al. (2006).[39] This means, patients with rituximab-naïve FL only, all patients had one or two prior lines of therapy (none had three or more), and all patients had Ann Arbor stage III-IV.

As mentioned previously, the most significant factor according to clinical experts to be considered for the MAIC of R^2 (AUGMENT) compared with R-CHOP (Van Oers) was prior rituximab exposure; but this was not included in the MAIC because all patients in the comparator study (Van Oers et al. (2006)) were rituximab-naïve. Another important factor noted by clinical experts was duration of prior response; this was also not included as a covariate in the MAIC. Therefore, the credibility of the MAIC is limited and results are not representative for the UK patient population. The company also concluded that the 100% rituximab-naïve population in Van Oers is not reflective of UK practice and used data from UK HMRN in the economic base-case analysis instead. Therefore, the indirect comparison using HMRN data will be critiqued next.

4.4.2 Indirect comparison of R^2 with R-CHOP/R-CVP based on HMRN data.

The company performed an indirect comparison with data from the Haematological Malignancy Research Network (HMRN). This included a comparison of R^2 with pooled data for R-CHOP/R-CVP for non-rituximab refractory patients.

'Due to small patient numbers for non-R-refractory patients receiving R-CHOP and R-CVP in the HMRN database, clinical expectation that R-CHOP and R-CVP would have similar efficacy in a relapsed/refractory setting and empirical data demonstrating this to be the case, efficacy analyses compared R^2 to the pooled R-CHOP/R-CVP' (CS, page 84).[1] As explained in section 4.4 of this report, the ERG does not think the company has presented any convincing evidence suggesting that R-CHOP and R-CVP have similar clinical effectiveness. The ERG believes the treatments are generally considered for different populations and their effectiveness is therefore difficult to compare.

There were 63 patients identified as receiving either R-CVP or R-CHOP as second- or later-line therapy. Comparisons were made for three time to event outcomes collected within the AUGMENT clinical study (OS, TTNLT and PFS). The definition of TTNLT as used for the HMRN analysis is time to documentation of new anti-lymphoma treatment from 'baseline'. The definition of PFS as used for the HMRN analysis is time from 'baseline' to disease progression (including transformation to diffuse large B-cell lymphoma) or death due to any cause and the definition of OS was time from start of treatment to date of death or if still alive censored at 18 December 2018.

The HMRN is a population-based cohort, established in 2004, comprising a total population of 3.8 million people covering the former adjacent UK Cancer Networks of Yorkshire and the Humber & Yorkshire Coast. The HMRN identified [redacted] patients who had received ≥1 prior line of chemotherapy for treatment of FL and were identified as being non-R-refractory or R-refractory after each treatment line. For the subgroup of patients who were non-R-refractory, [redacted] patients received R-CVP and [redacted] patients received R-CHOP as a second or later line therapy, although most patients ([redacted]%) received these treatments in second-line. Patients could be included in both treatment subgroups if they had received both treatments in different lines of therapy, for example, R-CHOP in second-line and R-CVP in third-line. The HMRN dataset only includes FL patients, not MZL patients (CS, Section B.3.3, page 134).[1]

The baseline characteristics that were commonly collected by the HMRN and the AUGMENT study are presented in Table 4.11.

Table 4.11: Covariates commonly collected across AUGMENT and HMRN datasets

Data source	HMRN	AUGMENT
Treatment	**R-CVP/R-CHOP (2L+ population)**	**R^2**
N	[redacted]	[redacted]
Age (years):		
Median	[redacted]	[redacted]
Range	[redacted]	[redacted]
n (%) Age >=60yrs	[redacted]	[redacted]
n (%) Age >=65yrs	[redacted]	[redacted]
Sex, n, %		
n (%) Males	[redacted]	[redacted]

Data source	HMRN	AUGMENT
n (%) Females	[redacted]	[redacted]
Number of prior systemic anti-lymphoma regimens:		
n (%) 1	[redacted]	[redacted]
n (%) 2	[redacted]	[redacted]
n (%) ≥ 3	[redacted]	[redacted]
Prior rituximab treatment, n (%)	[redacted]	[redacted]
POD24[a], n (%)	[redacted]	[redacted]
Fully Staged, n (%)	[redacted]	[redacted]
Bone marrow involved, n (%)	[redacted]	[redacted]
Nodal sites		
n (%) ≤4	[redacted]	[redacted]
n (%) >4	[redacted]	[redacted]
Bulky disease [b]	[redacted]	[redacted]
Stage		
n (%) I	[redacted]	[redacted]
n (%) II	[redacted]	[redacted]
n (%) III	[redacted]	[redacted]
n (%) IV	[redacted]	[redacted]

Source: CS, Appendix D3, Table 28, page 55.
2L+ = second or later line therapy; HMRN = Haematological Malignancy Research Network; NA = not applicable; R^2 = rituximab plus lenalidomide; R-CHOP = rituximab plus cyclophosphamide, doxorubicin, vincristine, prednisolone; R-CVP = rituximab plus cyclophosphamide vincristine prednisolone.
[a]) POD24 is defined as relapse within two years of initial chemoimmunotherapy.
[b]) Bulky disease has different definitions in AUGMENT and the HMRN dataset. AUGMENT: At least one lesion that is ≥ 7 cm or at least 3 lesions with 3 cm or larger in the longest diameter by investigator review. HMRN: At least one lesion that is ≥ 10 cm.

The same list of potential modifiers/prognostic variables discussed previously in the context of the ITC with published data, was used to identify the matching variables for this comparison. Therefore, Table 4.12 shows the same list of potential effect modifiers/prognostic variables (EM/PVs) that would ideally be adjusted for in a MAIC, as identified and validated by external clinical experts consulted by the company.[19]

Table 4.12: Potential EM/PVs that would ideally be adjusted for in a MAIC

Characteristic	highest priority	Included in MAIC	Comments
Previous exposure to rituximab	Yes	Yes	
FLIPI components: - Age (mean, or median if mean no reported, or % >60 years if neither reported) - Ann Arbor Stage (III-IV) - Nodal sites (>4) - High LDH	Yes	Yes Yes Yes	Included as: % Age ≥60yrs Not collected in HMRN
Refractory to last therapy	Yes	No	Not included in MAIC
Prior lines of therapy 1 vs. 2 vs. >2	Yes	Yes	
FLIPI risk group (low vs. intermediate vs. high)	Yes	No	Not collected in HMRN
FLIPI2+ components: - Serum beta-2 microglobulin high - Bone marrow involvement - Diameter of largest node >6 cm - Haemoglobin <12 dL/L			 Not included in MAIC Not included in MAIC Not included in MAIC Not included in MAIC
Time from last treatment			Not included in MAIC
POD24	Yes	Yes	
ECOG performance status (0–1 vs. 2+)			Not included in MAIC
Presence of B-symptoms			Not included in MAIC

Source: CS, Section B.2.9, pages 70-71.
EM = effect modifiers; ESS = effective sample size; FL = follicular lymphoma; FLIPI = Follicular Lymphoma International Prognostic Index; MAIC = matching-adjusted indirect comparison; MZL = marginal zone lymphoma; PV = prognostic variables; R = rituximab; R^2 = lenalidomide plus rituximab; R-CHOP = rituximab plus cyclophosphamide, doxorubicin, vincristine, prednisolone.
Notes: Adjusted N is the sum of the absolute weights. The patient characteristics presented are the potential EM/PVs that were included in the matching. The following potential EM/PVs had data for all included studies but were dropped from the matching to maintain a sufficiently large effective sample size for subsequent analysis: % previous rituximab exposure. The ESS and adjusted N including these variables were 4.2 and 0.1.

As can be seen from Table 4.12, matching was performed for the following variables:

- Age ≥60 years (FLIPI component)
- Ann Arbor Stage III-IV (FLIPI component)
- Nodal sites >4 (FLIPI component)
- Prior rituximab treatment
- Prior lines of therapy (1 vs. 2 vs. >2)
- POD24 status

The company stated that 'A key treatment effect modifier/prognostic factor that was not collected by the HMRN was the FLIPI risk category. However, three of the four FLIPI components were collected (only LDH was not collected)' (CS, pages 85-86).[1] Another key treatment effect modifier/prognostic

factor that was not included in the MAIC was 'refractory to last therapy', it is unclear why this factor was not included. In addition, all FLIPI2+ components (Serum beta-2 microglobulin high; bone marrow involvement; diameter of largest node >6 cm; and haemoglobin <12 dL/L), time from last treatment, ECOG performance status (0–1 vs. 2+), and presence of B-symptoms were not included in the MAIC.

Regarding ECOG performance status, the company states that ECOG PS 'was dropped from the MAICs because there were very few ECOG PS 2+ patients in AUGMENT/MAGNIFY, and the comparator studies also either had a small number of ECOG PS 2+ patients (hence were balanced) or did not report these data' (CS, page 71).[1] It was not reported whether ECOG PS was reported in the HMRN dataset.

The company was asked why 'sex' and 'bone marrow involved' were not included in the matching. The company responded that 'Sex was not identified as being a potential prognostic factor and/or treatment effect modifier in the list of variables that was validated by external clinical experts and was therefore not included as a matching variable' (Response to Clarification, Question A17a).[26] The company agreed that 'bone marrow involved' should have been considered as a matching variable given that it was identified as being a potential prognostic factor and/or treatment effect modifier.[26] In response to Question A17b, the company performed the comparison to R-CVP/R-CHOP with additional adjustment for bone marrow involvement, and concluded that the addition of this extra variable has had little impact on the results.[26]

In conclusion, several potential treatment effect modifiers/prognostic factors were not included in the MAIC; some because data were not reported in HMRN (FLIPI risk group, LDH,), some because the company regarded it not relevant (sex), and some for reasons that are not clear ('refractory to last therapy', all FLIPI2+ components (Serum beta-2 microglobulin high; bone marrow involvement; diameter of largest node >6 cm; and haemoglobin <12 dL/L), time from last treatment, ECOG performance status (0–1 vs. 2+), and presence of B-symptoms).

The main concerns are the same as for the previous MAIC (Section 4.1.1), i.e. the set of covariates included in the MAIC does not reflect the complete set of all possible covariates which affects the reliability of the OS and PFS results. Several covariates were not included in the MAIC because data were not reported in the HMRN dataset. Although this is through no fault of the company, it is a serious limitation which affects the reliability of the MAIC results.

As stated in Section 4.4.1 of this report, the credibility of the MAIC relies on the inclusion of all relevant treatment effect modifiers/prognostic factors. DSU report TSD 18[41] states that, 'An unanchored MAIC or STC effectively assumes that absolute outcomes can be predicted from the covariates; that is, it assumes that all effect modifiers and prognostic factors are accounted for. This assumption is very strong, and largely considered impossible to meet. Failure of this assumption leads to an unknown amount of bias in the unanchored estimate'.[41] As can be seen from the list of covariates included in the MAIC, it is clear that several treatment effect modifiers/prognostic factors were not included in the MAIC, including some that were considered key treatment effect modifiers/prognostic factors by the clinicians consulted by the company (FLIPI risk group, and 'refractory to last therapy').

The results of the matching for the EMs/PVs included in the matching are provided in Table 4.13 for all covariates included.

Table 4.13: Patient characteristics, observed and match-adjusted for the non-R-refractory population, comparing R² (AUGMENT) and R-CHOP/R-CVP (HMRN)

Characteristic	AUGMENT (R^2) (n=178)	HMRN (R-CHOP/ R-CVP) (n=63)	Adjusted R^2 (n=[redacted])
Patient characteristics			
% Prior rituximab	[redacted]	[redacted]	[redacted]
% Age ≥60yrs	[redacted]	[redacted]	[redacted]
% Ann Arbor stage III-IV	[redacted]	[redacted]	[redacted]
% Nodal sites ≤4	[redacted]	[redacted]	[redacted]
% 1 prior lines of therapy	[redacted]	[redacted]	[redacted]
% 2 prior lines of therapy	[redacted]	[redacted]	[redacted]
% Early relapse	[redacted]	[redacted]	[redacted]
Outcomes			
OS	Not estimable	63, [redacted] (NR)	NR
PFS (N, median (95% CI))	178, 39.4 months (NR)	63, [redacted] (NR)	NR

Source: CS, Appendix D3, Table 29, page 57.
EM = effect modifiers; ESS = effective sample size; FLIPI = Follicular Lymphoma International Prognostic Index; PV = prognostic variables; R = rituximab; R^2 = lenalidomide plus rituximab; R-CHOP = rituximab plus cyclophosphamide, doxorubicin, vincristine, prednisolone.

ERG comment: The comparison of R^2 versus R-CHOP/R-CVP was adjusted for the variables listed in Table 4.13, i.e. percentage of patients that had prior rituximab, age, Ann Arbor score and FLIPI score, nodal sites, prior lines of therapy and early relapse. The resulting population for the comparison of R^2 versus R-CHOP/R-CVP, are patients based on the baseline characteristics of patients in the HMRN dataset. As mentioned previously, two key treatment effect modifiers/prognostic factors (FLIPI risk group, and 'refractory to last therapy') were not included in the matching process. In addition, several covariates were not included in the MAIC because they were not reported in the HMRN dataset. Therefore, the credibility of the MAIC is limited.

Results of the MAIC are presented in Kaplan-Meier curves for OS, PFS, and TTNLT (CS, Figures 17-19, pages 88-90).[1] Hazard ratios (HRs) from the Cox Proportional-Hazard models comparing R^2 and R-CHOP/R-CVP are reproduced in Table 4.14. R^2 had a significant improvement in OS and TTNLT compared to R-CHOP/R-CVP and a benefit for TTNLT, but no evidence of a difference in PFS.

Table 4.14: Results from Cox Proportional Hazard models comparing R^2 and R-CVP/R-CHOP

Outcome	R^2, adjusted N	R-CHOP/R-CVP, N	HR (95% CI)[a]
OS	[redacted]	[redacted]	[redacted]
PFS	[redacted]	[redacted]	[redacted]
TTNLT	[redacted]	[redacted]	[redacted]

CI = confidence interval; HR = hazard ratio; N = number of patients; OS = overall survival; PFS = progression-free survival; R^2 = rituximab plus lenalidomide; R-CHOP; rituximab plus cyclophosphamide, doxorubicin, vincristine, prednisolone; R-CVP = rituximab plus cyclophosphamide, vincristine, prednisolone; TTNLT = time to next anti-lymphoma treatment.
[a]) bootstrapped CI.

The company should have presented crude unadjusted differences alongside the MAIC estimates, in line with the recommendations in NICE Decision Support Unit (DSU) technical support document (TSD) 18[41] to enable comparisons between the adjusted MAIC and unadjusted results. No such estimates have been presented, apart from the Kaplan-Meier curves in Figures 17-19 (CS, pages 88-90).[1]

NICE DSU TSD report 18 lists several themes that should be considered and addressed explicitly when reporting population-adjusted analyses (See TSD 18, pages 64-65).[41] In Appendix 2 these themes are reproduced with an ERG comment how they were addressed in this submission. As can be seen from Appendix 2 not all themes were addressed in the CS.

In conclusion the results of the MAIC should be treated with a high degree of caution due to the fact that potentially important covariates were excluded from the matching models, small sample sizes and assumptions about the equivalence of R-CHOP and R-CVP in the HMRN data and differences in PFS definitions and length of follow-up between the two data sources, The analysis also used an unanchored MAIC involving two single treatment arms from different studies, as there was no relevant comparative trial data. This analysis makes the assumption that all effect modifiers and prognostic factors are accounted for in the model, which in practice is difficult to achieve as, in this case, one or both studies do not measure a specific variable.

4.5 Additional work on clinical effectiveness undertaken by the ERG

No further additional work on clinical effectiveness was undertaken by the ERG.

4.6 Conclusions of the clinical effectiveness section

The company submission and response to clarification provided sufficient details for the ERG to appraise the literature searches. A good range of resources were searched and the searches were transparent and reproducible. One set of searches was conducted to identify both efficacy and safety evidence. Separate searches were conducted to identify cost effectiveness studies, health-related quality of life studies, and healthcare resource use data.

The company submission included six studies that were deemed relevant by the company. Four studies evaluated R^2, one of these was an RCT of R^2 versus R-monotherapy (the AUGMENT trial[35]), the other three[36-38] did not include relevant comparators according to the NICE scope. The remaining two studies evaluated R-CHOP versus CHOP (Van Oers et al., 2006[39]) and O-Benda versus bendamustine monotherapy (the GADOLIN trial[40]). The trial by Van Oers et al. (2006)[39] was used by the company for an unanchored indirect comparison (using individual arms of different studies) of R^2 versus R-CHOP. However, the study only included rituximab-naïve patients and was therefore not representative for the UK patient population. The GADOLIN study[40] was used by the company for an unanchored indirect comparison of R^2 with O-Benda. However, as explained in Sections 3.3 and 4.4 of this report, O-Benda is not considered by NICE to be a relevant comparator for this appraisal; therefore, this study was ignored in this report.

In conclusion, the CS included one relevant study, for the comparison of R^2 versus R-monotherapy: the AUGMENT trial.[35] All patients in this trial were non-R-refractory. In addition, the company performed an unanchored indirect comparison of R^2 versus R-CHOP and R-CVP, using data for R^2 from the AUGMENT trial and pooled data for R-CHOP/R-CVP from the HMRN database.

The AUGMENT trial[35] is a randomised, double-blind, multicentre, Phase III study of R^2 versus rituximab plus placebo (R-mono) in non-R-refractory patients with FL Grade 1–3a or MZL. The study was conducted across 96 sites in 17 countries. The trial did not include any patients from the UK. The primary efficacy analyses were conducted on the ITT population, defined as all randomised patients. The primary endpoint of the study was PFS, as assessed by the Independent Review Committee (IRC).

Results from the AUGMENT trial show favourable results for R^2 when compared to R-mono in terms PFS with a greater median PFS ([redacted] vs. [redacted] months; HR [redacted] (95% CI: [redacted]). However, there was no evidence of a difference in OS with a HR of 0.61 (95% CI: 0.33 to 1.13) for patients treated

with R^2 compared to R-mono. At the time of the analysis the OS data was immature with 16 deaths on R^2 and 26 deaths on R-mono at the time of the analysis. Overall response rate (ORR) was significantly greater for R^2 compared with R-mono (78% vs. 53%; p<0.0001). The complete response (CR) rate was also greater for the R^2 arm compared with R-mono (34% vs. 18%; p=0.001). Results for R^2 versus R-mono in MZL patients were generally less favourable for R^2 than in FL patients. However, it is important to note that PFS outcomes in the MZL subgroup are difficult to interpret because of the small sample size (63 patients in total) and imbalance in baseline prognostic factors. In terms of health-related quality of life, no clinically meaningful change from baseline in the GHS/QoL domain of the QLQ-C30 was observed across any of the post-baseline assessment visits, regardless of treatment group. Between-group differences in mean changes were small and not clinically meaningful across all assessment visits and did not differ between FL and MZL patients.

R^2 was associated with more grade 3-4 TEAEs and SAEs when compared to R-mono, especially lenalidomide/placebo related adverse events; but rituximab-related grade 3-4 TEAEs and SAEs were also more frequent in the R^2 arm than in the R-mono arm. R^2 was also associated with more TEAEs leading to dose reductions, dose interruptions and discontinuations of lenalidomide/placebo or rituximab when compared to R-mono. Adverse events are generally the same for FL and MZL patients; however, AEs for MZL patients are based on small numbers.

The company performed three unanchored indirect comparisons, two using data from published evidence and one using data from HMRN:

- R^2 versus R-CHOP for non-rituximab refractory patients, based on comparator data from a study by Van Oers et al. 2006[39] comparing R-CHOP with CHOP (only the R-CHOP arm was used in the analyses).
- R^2 versus established clinical management without lenalidomide, i.e. O-Benda for rituximab refractory patients, based on comparator data from a study by Sehn et al. 2016[40] comparing O-Benda with bendamustine monotherapy (only the O-Benda arm was used in the analyses).
- R^2 versus pooled data for R-CHOP/R-CVP for non-rituximab refractory patients using data from HMRN.

As mentioned above, the two unanchored indirect comparisons using published evidence have been ignored in this report. R^2 versus R-CHOP, because the study by Van Oers is not representative for UK patients, and R^2 versus O-Benda because O-Benda is not a relevant comparison for this appraisal according to NICE.

Results from the remaining MAIC (R^2 versus pooled data for R-CHOP/R-CVP for non-rituximab refractory patients using data from HMRN) show a significant improvement in OS (HR = [redacted] (95% CI: [redacted]) and TTNLT (HR = [redacted] (95% CI: [redacted]) compared to R-CHOP/R-CVP, but no evidence of a difference in PFS (HR = [redacted] (95% CI: [redacted]).

The results of the MAIC should be treated with a high degree of caution due to the fact that potentially important covariates were excluded from the matching models, small sample sizes and assumptions about the equivalence of R-CHOP and R-CVP in the HMRN data and differences in PFS definitions and length of follow-up between the two data sources, The analysis also used an unanchored MAIC involving two single treatment arms from different studies, as there was no relevant comparative trial data. This analysis makes the assumption that all effect modifiers and prognostic factors are accounted for in the model, which in practice is difficult to achieve as, in this case, one or both studies do not measure a specific variable.

5. COST EFFECTIVENESS

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

The company conducted searches for cost effectiveness, health-related quality of life and healthcare resource use evidence. A good range of databases, conference proceedings and additional resources were searched. The company submission and clarification response provided sufficient detail for the ERG to be able to appraise the searches conducted by the company.

5.1.1 Searches performed for cost effectiveness section

The following paragraphs contain summaries and critiques of all searches related to cost effectiveness presented in the company submission.

Appendices G, H and I of the CS report the literature searches used to identify cost effectiveness, health-related quality of life and healthcare resource use studies. Separate sets of searches were run for each. Searches were conducted on 8 February 2019. A summary of the sources searched is provided in Table 5.1. The CS reported that targeted literature searches were conducted to identify adverse event disutility values, FL and MZL prognosis studies, and data on response rates, OS and PFS: these targeted searches were not provided. The company described how these data were identified via targeted literature searches in their response to the ERG clarification letter.

Table 5.1: Resources for the cost effectiveness, health-related quality of life and healthcare resource use literature searches

Search strategy element	Resource	Host/Source	Date Range	Date Searched
Electronic databases	MEDLINE	Embase.com	Not reported	8 February 2019
	Embase			
	MEDLINE In-Process	PubMed	Not reported	8 February 2019
	EconLit	EBSCO	Not reported	8 February 2019
	NHS EED	CRD interface	Not reported	8 February 2019
	HTA		Not reported	8 February 2019
Conference proceedings	ISPOR International	http://www.ispor.org/heor-resources/presentations-database/search	2017, 2018	February 2019
	ISPOR European	http://www.ispor.org/heor-resources/presentations-database/search	2017, 2018	February 2019
	ASH	http://www.hematology.org/Annual-Meeting/Archive.aspx	2017, 2018	February 2019
	EHA	https://ehaweb.org/congress/previous-congresses/	2017, 2018	February 2019
	ICML	http://www.lymphcon.ch/icml/website/icml-abstracts-books/icml-abstract-books-1981-2011.html	2015, 2017	February 2019

Search strategy element	Resource	Host/Source	Date Range	Date Searched
	ASCO	https://meetinglibrary.asco.org/browse-meetings/	2017, 2018	February 2019
HTA Agencies	NICE	https://www.nice.org.uk/		February 2019
	SMC	https://www.scottishmedicines.org.uk/		February 2019
	AWMSG	http://www.awmsg.org/		February 2019
	HAS	https://www.has-sante.fr/portail/jcms/r_1455081/en/home-page?portal=r_1455081		February 2019
	SLV	https://legemiddelverket.no/English https://www.legemiddelsok.no/		February 2019
Bibliographic searches of key systematic review and meta-analysis articles were conducted to ensure that initial searches captured all the relevant economic studies				

HTA = Health Technology Assessment Database; NHS EED = NHS Economic Evaluation Database; ISPOR = International Society for Pharmacoeconomics and Outcomes Research; ASH = American Society of Hematology; EHA = European Hematology Association; ICML = International Conference on Malignant Lymphoma; ASCO = American Society of Clinical Oncology; NICE = National Institute for Health and Care Excellence; SMC = Scottish Medicine Consortium; AWMSG = All Wales Medicines Strategy Group; HAS = The Haute Autorité de Santé; SLV = Statens legemiddelverk.

ERG comment:

- MEDLINE and Embase were searched simultaneously using embase.com. This approach is not recommended. A simultaneous multi-file search such as this should include both MeSH and EMTREE subject headings to ensure that all subject indexing terms are searched; however, all of the economic search strategies only included EMTREE terms which may have impaired how well the strategies performed.
- There were no details about which MEDLINE segments were searched (Table 35, Table 44 and Table 54 in Appendix G of the CS).[34]
- Date ranges were not reported for any of the economic related database searches.
- The CS reported that MEDLINE In-Process was searched using PubMed (Table 36, Table 45 and Table 55). This is inaccurate, as the search limit used in PubMed identifies 'Ahead of print' and recently added records, not in-process records: (publisher[sb] NOT pubstatusnihms NOT pubstatuspmcsd NOT pmcbook) OR (pubstatusaheadofprint). Therefore in-process records were actually excluded from the company's PubMed search.
- The company reported searching NHS EED and the HTA database via the Cochrane Library using the CRD search interface. This is incorrect as NHS EED and HTA are no longer available on the Cochrane Library or have anything to do with Cochrane. The company conducted the NHS EED and HTA searches via the Centre for Reviews and Dissemination (CRD) interface, and misreported using the Cochrane Library.
- Truncation and proximity operators were used more often in the cost effectiveness searches than in the clinical effectiveness searches. As with the clinical effectiveness searches, there were too few synonyms. However, the 'syn' operator was included, and embase.com enables automatic synonym searches when this operator is added to an EMTREE term. The ERG does not have access to Embase.com to test the impact of this on search performance.
- The search strategies used in MEDLINE In-Process (PubMed), EconLit, and NHS EED/HTA only included a population facet of search terms, and so were sensitive enough to identify studies for all of the economic sections (cost effectiveness, health-related quality of life and healthcare resource

use). The embase.com search strategies included an additional facet of search terms for each of the economic sections (cost effectiveness, health-related quality of life and healthcare resource use); three separate searches were conducted in embase.com.

- It is not clear if the search facets used to identify cost effectiveness, health-related quality of life and healthcare resource use were based on validated search filters, such as those published on the ISSG Search Filters Resource website: https://sites.google.com/a/york.ac.uk/issg-search-filters-resource/
- A good range of conference proceedings and HTA organisation websites were searched, and although full details of these searches were not provided in the CS, they were provided in response to the ERG clarification letter.
- Targeted literature reviews were referred to in the CS, but no details were reported. In response to the ERG clarification letter the company provided details of the targeted literature reviews, and how adverse event disutility values, FL and MZL prognosis studies, and data on response rates, OS and PFS were identified. Data were identified by investigating the clinical systematic literature review results, reviewing previous NICE technology appraisals, and a targeted literature search of PubMed.

5.1.2 Inclusion/exclusion criteria used in the study selection

In- and exclusion criteria for the review on cost effectiveness studies, utilities and costs and resource use are presented in Table 39 of Appendix G, Table 48 of appendix H, and Table 58 of Appendix I of the CS, repectively.[34]

ERG comment: The ERG agrees that the eligibility criteria are suitable to fulfil the company's objective to identify cost effectiveness studies.

5.1.3 Included/excluded studies in the cost effectiveness review

In total, 24 cost effectiveness studies met the pre-defined eligibility criteria.[13, 20, 22, 23, 43-62] These were extracted from 31 publications of which 22 full publications and nine HTA submissions. Details of these studies were provided in Tables 23 and 24 of the CS. The search for utility studies resulted in 38 included studies, for which details and references were provided in Table 49 of Appendix H of the CS.[34] The search for costs and resource use resulted in 17 included studies, for which details and references were provided in Table of Appendix I of the CS.[34]

ERG comment: The rationales for excluding CE studies after full paper reviewing are considered appropriate given the defined in- and exclusion criteria.

5.1.4 Conclusions of the cost effectiveness review

The CS provides an overview of the included cost effectiveness, utility and resource use and costs studies, but no specific conclusion was formulated.

ERG comment: Eligibility criteria were suitable for the SLR performed. However, it was not fully clear to the ERG how the information obtained from the SLR was implemented in the de novo analysis. For instance, the company stated in B.3.1 of the CS that they had identified four economic evaluations that had a UK perspective and were of potential value to inform this submission. They then stated that 'more details of how these evaluations have informed the de novo analysis are discussed in Section B.3.2.'.[1] However, Section B.3.2. of the CS does not contain any information on the use of these evaluations.

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

Table 5.2: Summary of the company's economic evaluation (with signposts to CS)

	Approach	Source/Justification	Signpost (location in CS)
Model	Partitioned survival model	Makes use of the PFS and OS data directly, ensuring that estimated survival outcomes versus observed outcomes are matched.	B.3.2
States and events	Progression-free, post-progression, death		B.3.2
Comparators	Non-rituximab-refractory patients: R-CHOP and R-CVP rituximab-refractory patients: O-Benda	Expert opinion	B.3.2
Population	The patient population considered in the model is, adult patients with previously treated FL or MZL (pooled). The model is split into two subpopulations: non-rituximab-refractory and rituximab-refractory patients.	In line with the proposed licence. FL and MZL populations were pooled due to the similar prognosis of FL and MZL patients, and the difficulty in sourcing MZL-specific data	B.3.2
Treatment effectiveness	Non-rituximab-refractory: Unanchored MAIC using AUGMENT and HMRN Rituximab-refractory: Unanchored MAIC using MAGNIFY and GADOLIN		B.3.3
Adverse events	Grade 3 and 4 based on trial data		B.3.3
Health related QoL	EQ-5D-3L data from AUGMENT	NICE reference case	B.3.4
Resource utilisation and costs	NHS and Personal Social Services	NICE reference case	B.3.2
Discount rates	3.5% discount rate was used for utilities and costs	NICE reference case	B.3.2.
Subgroups	non-rituximab-refractory and rituximab-refractory patients		B.3.9
Sensitivity analysis	Probabilistic and deterministic sensitivity analyses and scenario analyses	NICE reference case	B.3.8

FL, follicular lymphoma; MZL, marginal zone lymphoma; R-CHOP, rituximab plus cyclophosphamide, doxorubicin, vincristine, prednisolone; R-CVP, rituximab plus cyclophosphamide vincristine prednisolone; O-Benda, obinutuzumab plus bendamustine; MAIC, matching-adjusted indirect comparison; HMRN, Haematological Malignancy Research Network.

5.2.1 NICE reference case checklist (TABLE ONLY)

Table 5.3: NICE reference case checklist

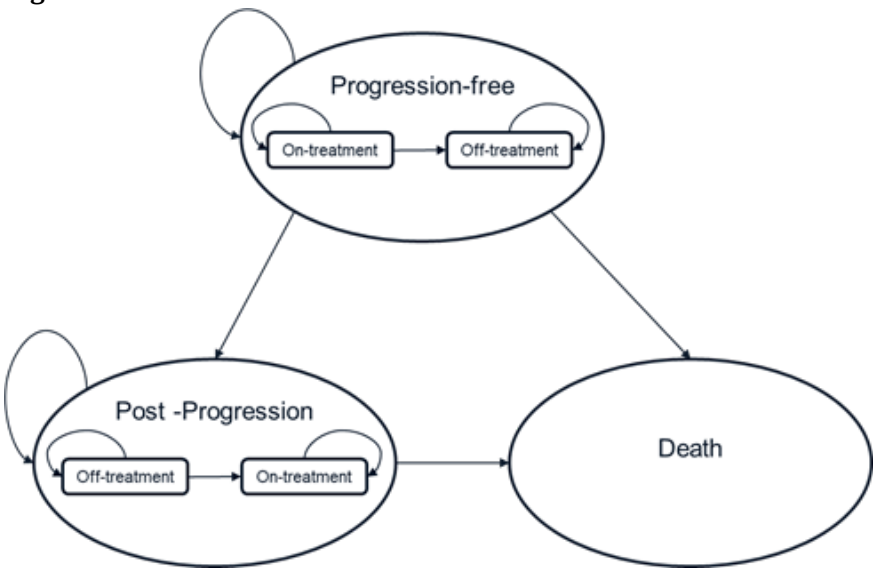
Elements of the economic evaluation	Reference Case	Included in submission	Comment on whether *de novo* evaluation meets requirements of NICE reference case
Population	As per NICE scope	Yes, although divided in non-rituximab-refractory and rituximab-refractory patients	
Comparator(s)	Therapies routinely used in the National Health Service (NHS), including technologies regarded as current best practice	Non-rituximab-refractory patients: R-CHOP and R-CVP Rituximab-refractory patients: O-Benda	R-mono was not included in the evaluation while it was listed in the scope. The company added a comparison of R^2 and R-mono in response to clarification questions. NICE have explicitly stated that O-Benda is not considered a relevant comparator for disease that is refractory to rituximab. The ERG report does not contain information on this comparator.
Type of economic evaluation	Cost effectiveness analysis	Yes	
Perspective on costs	NHS and Personal Social Services (PSS)	Yes	
Perspective on outcomes	All health effects on individuals	Yes	
Time horizon	Sufficient to capture differences in costs and outcomes	Yes	
Synthesis of evidence in outcomes	Systematic review (SLR)	Yes	
Measure of health effects	Quality adjusted life years (QALYs)	Yes	
Source of data for measurement HRQoL	Described using a standardised and validated instrument	Yes	
Source of preference data for valuation of	Time-trade off or standard gamble	Yes	

Elements of the economic evaluation	Reference Case	Included in submission	Comment on whether *de novo* evaluation meets requirements of NICE reference case
changes in HRQoL			
Discount rate	An annual rate of 3.5% on both costs and health effects	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 modelling	Yes	
NHS = National Health Service; NICE = National Institute for Health and Care Excellence; PSS = Personal Social Services; QALY = quality-adjusted life year; SLR = systematic literature review			

5.2.2 Model structure

A cohort-level partitioned survival model (PSM) was developed with three health states: progression-free (PF), post-progression (PP) and death. The company argued that a PSM was more appropriate than a state transition model (STM) because of a lack of data on post progression survival (PPS). According to the company, the relevant comparators for this submission are not included in the head-to-head study with R^2 (AUGMENT); therefore, the data available for informing PPS for the comparators are reduced to available published data or alternative sources. All patients start 'on treatment' in the PF health state. Subsequently, patients either remain on treatment or come off treatment before progressing or dying per cycle. Within PP, patients can have a treatment-free interval before receiving subsequent therapy. Patients in the PP on treatment health state remain in this health state until they die. The model was programmed in Microsoft Excel.

Figure 5.1: Model structure



Source: Based on Figure 21 of the CS

ERG comment: The main concern of the ERG relates to the use of a PSM instead of a STM. The use of a PSM instead of a STM was justified by the lack of data of relevant comparators in the head-to-head study with R^2 to inform a state transition model. Despite the potential limitations of a state transition model, a partitioned survival analysis has several limitations related to the extrapolation (as mentioned in NICE DSU TSD 19[63]). The ERG requested a scenario analysis using a STM as a scenario, as recommended in TSD 19, which the company did not deliver. The company argued that because of the weight of the limitations in the STM approach, combined with the specifics of the data available for this decision problem, constructing a state transition model is not applicable for this submission. The ERG acknowledges that every model approach has its limitations, and that the lack of data for the R-CHOP and R-CVP posed a problem populating a STM. However, the lack of a structural link between endpoints in a PSM may lead to biased extrapolations.[63] Therefore, according to the ERG, and in line with recommendations from TSD 19, STM should be used alongside PSM to assess the plausibility of extrapolations, if only for the comparison in the pivotal trial.

5.2.3 Population

R^2 does not currently have a UK marketing authorisation. The patient population considered in the model is in line with the proposed license: adult patients with previously treated FL or MZL. Due to the similar prognosis of FL and MZL patients, and the difficulty in sourcing MZL-specific data, FL and MZL populations were pooled throughout the economic analysis. Non-rituximab refractory patients and rituximab refractory patients were modelled separately because the company assumed the relevant comparators for these patients would be different. The patient cohort considered in the model varies per population. The patient starting age and gender were matched to the data source used for the comparator arms (for non-R refractory patients this was the HMRN: mean age [redacted] years, percentage female [redacted]). Body surface area (BSA) data were taken from individual patients in the AUGMENT study (mean BSA 1.85 m^2).

ERG comment: The main concern of the ERG relates to pooling the FL and MZL populations throughout the economic analysis. In response to clarification question B1 the company provided an overview of the population the evidence in the economic analysis was based on. All evidence of the comparators was based on datasets that only contained patient with FL, while the AUGMENT trial

contained patients with FL and MZL. The AUGMENT trial was used as the source for utilities for R^2 as well as the comparators, and as the source of subsequent treatments for R^2. Furthermore, the company provided exploratory post-hoc analyses which investigated the impact of the histology (MZL/FL) on the outcome of PFS in the AUGMENT trial data to justify that the prognosis and comparative effectiveness are similar for FL and MZL. These analyses showed that neither the interaction term between the randomised treatment arm and histology, nor histology were statistically significant (p-value >0.05). The company argued that clinicians during the expert meeting stated resource use for FL and MZL patients was similar. Analysis of AUGMENT quality of life data showed that if histology was included in the mixed effects regression model used for the utilities, this resulted in a mean utility difference of 0.03 for MZL patients, however this was not statistically significant (p=0.145). The company provided an FL-only scenario analysis (discussed in section 5.2.11).

5.2.4 Interventions and comparators

R^2 does not currently have a UK marketing authorisation. The R^2 dosing regimen within the model is lenalidomide 20 mg orally once daily on days 1–21 of repeated 28-day cycles for up to 12 cycles of treatment. Rituximab is given as 375 mg/m2 every week in Cycle 1 (days 1, 8, 15 and 22) and Day 1 of every 28-day cycle for Cycles 2–5. This is in line with the recommended dose in the SmPC.[18] Patients with moderate renal impairment start on a dose of 10 mg of lenalidomide if CrCl is ≥30 ml/min but <60 ml/min. These criteria were met by [redacted]% of patients in AUGMENT and [redacted]% in MAGNIFY (R-refractory population), and these proportions are used to inform the starting dose in the model for the non-R-refractory and R-refractory populations, respectively.

In AUGMENT R^2 is compared to R-mono. The company states that according to clinical experts, R-mono is rarely used in the relapsed/refractory setting in UK clinical practice.[31] Instead, comparators for R^2 in the non-R-refractory population are rituximab in combination with chemotherapy; predominantly R-CHOP and R-CVP. Experts also stated that R-Benda is primarily used in a first-line setting and clinicians are reluctant to re-challenge relapsed/refractory patients with bendamustine in subsequent lines of therapy.[19] Therefore, R-mono and R-Benda were not considered relevant comparators for the non-R-refractory population. For the R-refractory population the company states that clinical experts believe that O-Benda has largely replaced use of bendamustine.[19]

ERG comment: The main concerns of the ERG relate to: a) the inclusion of O-Benda as a comparator while NICE have explicitly stated it is not considered a relevant comparator for disease that is refractory to rituximab, b) omitting R-mono as a comparator (based on expert opinion) although listed in the scope and given the direct evidence available.

a) The ERG did not include O-Benda in her review as NICE has explicitly stated it is not considered a relevant comparator for disease that is R-refractory.

b) In response to question B3 the company provided an analysis of R^2 versus R-mono based on the AUGMENT trial data.

5.2.5 Perspective, time horizon and discounting

The analysis took an NHS and Personal Social Services (PSS) perspective. Discount rates of 3.5% are applied to both costs and benefits. The model cycle length of 28 days with a 40-year time horizon and half cycle-correction is applied.

ERG comment: In the CS, the company stated that a 40-year time horizon was used. The model output showed this was in fact a life time horizon.

5.2.6 Treatment effectiveness and extrapolation

The main source of evidence on treatment effectiveness used for intervention and comparators was the AUGMENT study[64] for R^2 and HMRN data[65] for R-CHOP and R-CVP. The AUGMENT study is a Phase III, multicentre, double-blind, randomised study comparing R^2 versus R-mono in patients with non-R-refractory/relapsed FL or MZL. Only data from the R^2 arm and from the 22 June 2018 data cut-off were used in the model. The HMRN is a population-based cohort covering the Yorkshire and Humber & Yorkshire Cancer Networks for all patients newly diagnosed with a haematological malignancy between 2004 and 2016. No data on MZL patients was available in the HMRN.

The Phase III study by van Oers et al. (2006)[39] on R-CHOP was not used in the base-case analysis because all patients were R-naïve, which was not thought to be reflective of current clinical practice in the UK. Also, with prior rituximab exposure being an important effect modifier, matching with the R^2 arm of AUGMENT data would be hampered. The van Oers et al. study data were used in a scenario analysis. For R-CVP, no trial-based evidence was found.

As the company considered OS and PFS in HMRN to be similar between R-CHOP and R-CVP, and clinical opinion suggested that in the relapsed/refractory setting it would not be unreasonable to assume the efficacy of R-CHOP and R-CVP to be similar, HMRN data for R-CHOP and R-CVP were pooled. Data from AUGMENT (n=103) were then matched to the pooled data from HMRN for R-CHOP and R-CVP (n=63). For the economic model, this implied that the comparisons of R^2 vs. R-CHOP and R-CVP had identical outcomes for effectiveness (LYs and QALYs) and only differed with respect to costs.

Parametric survival curves were fitted to the matched patient level data from AUGMENT and HRMN and were then used to extrapolate survival beyond study follow-up. Survival analysis was performed for OS, PFS, TTNLT, and ToT (time on treatment). The CS mentioned four criteria for selection of the curves: 1) proportional hazards assumption based on log cumulative hazard plots 2) visual inspection, Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC) 3) clinical plausibility, and 4) implausible curve crossings (e.g. OS moving below TTNLT) before 15 years of follow-up.

PFS and ToT data were used to determine the number of patients staying in the PF (on and off treatment) health states. PFS, TTNLT and OS data were used to determine the number of patients transitioning to the PP (on and off treatment) health states. The number of patients transitioning to the death state was derived using OS data.

The curves were adjusted for treatment waning, which was assumed to occur at five years. After this time point, the comparator hazard of progressing or dying was applied to the R^2 arm. This five-year time point was selected for the base-case as the company stated it to be consistent with previous NICE submissions in the same disease area (TA472[20] and TA137[59]).

Overall survival

As the log-cumulative hazard versus time plot for OS suggested that the proportional hazards assumption was violated, stratified models were used. Although AIC and BIC indicated that the exponential distribution fitted best on average, the Weibull distribution was selected for the base-case analysis. The company did not explain why the exponential distribution was not used, but stated that AIC/BIC for Weibull suggested a reasonable fit, and Weibull was also used in TA137[59]. The curves were adjusted for general population mortality (age and gender matched) so overall survival in the model would not exceed survival in the general population.

For the R-mono comparison which was added upon request of the ERG in the response to clarification, the company chose Weibull for both arms again. No rationale or diagnostic plots were provided but statistical fit based on AIC/BIC was best for exponential, like in the R-CHOP/R-CVP comparisons.

Progression free survival

As the KM curves for R^2 and R-CHOP/R-CVP at first slightly diverge but then converge and even overlap, this was suggestive of a non-constant treatment effect. This was confirmed by the log-cumulative hazard plot, which was non-parallel. The company then decided to model the PFS for R^2 using the KM data until the maximum follow-up of 46.7 months, and applied the comparator hazard to extrapolate further. In this way, the company stated in the CS, the relative treatment effect of R^2 vs. R-CHOP/R-CVP based on the MAIC was accurately reflected. Parametric curves were still fitted to each arm to be able to test assumptions used in the base-case. Based on the AIC/BIC, the Weibull distribution makes the best fit to the R-CHOP/R-CVP data whereas the exponential and log-logistic distributions seem to fit the R^2 data best. Nevertheless, the company chose to use the generalised gamma curve in the base-case, because the Weibull would cross the TTNLT curve in R-CHOP/R-CVP at approximately eight years, which would be clinically implausible since it would be unlikely that patients have their next treatment prior to progression in clinical practice.

Finally, the curves were adjusted to ensure that long-term PFS estimates would not be higher than TTNLT or OS.

For the R-mono comparison, a simpler approach was taken, using log-logistic for both arms.

Time to next anti-lymphoma treatment

From the cumulative hazard plot, proportional hazards seemed reasonable but not definitive, and therefore stratified models were used, with unstratified models explored in a scenario. Based on the AIC/BIC, the exponential distribution best fitted the R^2 data, and the log-normal distribution fitted best to R-CHOP/R-CVP. However, as the exponential distribution would result in crossing of PFS and TTNLT around seven years, the company chose the log-normal distribution for the base-case analysis for both arms.

Finally, in line with what was done for PFS, the curves were adjusted to ensure they would not be higher than OS.

For the R-mono comparison, the generalised gamma was used for both arms. AIC and BIC were provided in the model for a series of distributions, but the choice for generalised gamma was not further justified.

Time on treatment

ToT data were used to determine the proportion of patients on treatment to calculate overall drug costs. Parametric survival curves were fitted to the ToT data which, however, produced a poor fit. Therefore, the company chose to use the KM data directly in the model, and maximum treatment durations were used to cap ToT

For the R-mono comparison the same approach was used, that is, KM data were used.

ERG comment: The main concerns of the ERG relate to: a) the uncertainty introduced by the indirect comparison of R^2 with R-CHOP and R-CVP based on only 63 patients - which seems to be underlined by b) the counterintuitive results for the R-mono comparison c) the lack of justification for the choice of time-point at which treatment effect ends d) the seemingly arbitrary way of selecting the curves used

for extrapolating, which seems mainly guided by trying to avoid implausible curve crossings e) in particular the choice of the OS curves and f) the PFS curves, but also g) TTNLT curves. In addition, h) an error was found when running the scenario using van Oers data for efficacy.

a) The ERG has serious doubts about how trustworthy the results of the indirect treatment comparisons with R-CHOP and R-CVP are, given that HMRN data are based on only 63 patients in total, were collected much earlier than the data from AUGMENT, and consist of two treatment regimens which may not be as similar as assumed. Although the ERG appreciates that R-CHOP and R-CVP data were pooled to obtain a larger sample size, it is still small and the pooling may have introduced additional bias as the KM curves from the HMRN report[65] show a rather consistent difference in favour of R-CHOP, which may be a result of the fact that the target population for R-CHOP is the younger and fitter group, enhancing efficacy. Furthermore, data collection for patients in HMRN started much earlier (from 2004 onwards), and a time effect interfering with the treatment effect cannot be ruled out, given the continuous improvements in clinical practice. These changes in clinical practice may be illustrated by the fact that in the modelled subsequent therapies, the proportion of targeted therapies was 0% for the HMRN R-chemo cohort and 6.7% for the R^2 arm in AUGMENT. The uncertainty associated with the indirect comparison was not captured in the model and as such cannot be quantified but its impact is likely substantial.

b) In their response to clarification (question A7b), the company stated that R-CHOP and R-CVP are considered more effective than R-mono.[26] One would expect the model to confirm this. However, in the additional analysis that the company provided upon request of the ERG, the ICER of R^2 versus R-mono was substantially higher at £22,580 vs £11,471 for R^2 versus R-CHOP. This was predominantly caused by the fact that LYs and QALYs for R^2 were lower, while costs were higher. So, when using data from the direct comparison as per AUGMENT, R^2 was more costly and less effective than when using results from the MAIC. This again raises the question whether the indirect comparison provided valid results, as the MAIC seems to inflate efficacy and lower costs for R^2. The model does not accommodate quantification of this uncertainty and so the ERG cannot provide an estimate of its potential impact.

c) The company assumed treatment waning to start at five years, based on previous STAs. Upon the ERG's request in the clarification phase to further justify this choice of timepoint, the company replied that neither TA472[20] or TA137[59] appeared to present evidence to support their assumptions, even though treatment effect was a key uncertainty in these appraisals, having a large impact on the ICERs. The company further argues that five years is considered conservative as the immunomodulatory effect of lenalidomide could promote a longer treatment effect versus R-chemo's. The company also argued that choice of time point did not have a huge impact on the results when tested at three or 10 years. The ERG considers the company's choice of time point to be rather arbitrary and a shorter or longer duration of treatment effectiveness may be equally likely. As in the company base-case, the ERG ran scenarios varying the time point to three and seven years.

d) The company proposed a systematic way of selecting the parametric curves for extrapolating, consisting of four steps i.e. 1) proportional hazards assumption based on log cumulative hazard plots 2) visual inspection, AIC and BIC 3) clinical plausibility, and 4) implausible curve crossings (e.g. OS moving below TTNLT) before 15 years of follow-up. In the actual selection, however, it is difficult to see how these criteria were handled. For OS, PFS, and TTNLT, the CS states that 'all curves fit the data reasonably well'. Avoiding implausible curve crossing seemed to be the main argument for selection.

e) For OS, the company's argument for choosing the Weibull distribution over the better fitting exponential distribution was that the AIC/BIC for Weibull suggested a reasonable fit, and Weibull was also used in TA137[59] on R-mono. Given the company's claim that R^2 is essentially different from R-mono, the ERG is not convinced that OS in R^2 would be logically comparable to OS in TA137. Given the criteria that the company stated to have taken into account for selecting the curves, it is not clear to the ERG how the Weibull could be preferred over the exponential distribution. The ERG base-case used the exponential distribution for both arms. For the R-mono comparison the same argument applied and so the ERG base-case also incorporated the exponential distribution for OS.

f) The company interpreted the slight divergence and subsequent convergence/overlap of the KM curves for R^2 and R-CHOP/R-CVP as a non-constant treatment effect. They then decided to model PFS for R^2 using the KM data until the maximum follow-up after which the comparator hazard was applied to extrapolate further. In this way, the CS stated, the relative treatment effect of R^2 vs. R-CHOP/R-CVP based on the MAIC was accurately reflected. The ERG fails to see why and how this way of modelling PFS would accurately reflect the, as the company stated "non-constant", relative treatment effect. The overlap of the KM curves may as well be indicative of the absence of a treatment effect. At the end of follow-up, the KM for R^2 is higher than any of the estimated survival curves. See Figure 5.2 with the parametric PFS curves alongside the KM data + comparator hazard that was actually used in the base-case. The ERG considers this approach to favour R^2 even though the parametric PFS curves for R-CHOP/R-CVP are mostly higher than those for R^2, in particular from the five-year point onwards. Also, choosing the point of last follow-up as a starting point to extrapolate further is quite arbitrary. At any other timepoint, the start of the extrapolation could have been substantially different. Furthermore, near the tail of the KM-curve the number of patients approaches zero (exact numbers difficult to see from the CS) which would increase the uncertainty of the extrapolation that follows.

   For R-CHOP/R-CVP, the generalised gamma that was used in the company base-case appears to underestimate PFS in the first year (Figure 5.3) and remains lower than most of the other non-parametric curves. Given that generalised gamma does not provide the best statistical fit, the ERG considers this a sub-optimal choice. As the company advocated separate model types for the two treatment arms, which seems reasonable, the ERG base-case includes the log-logistic curve for R^2 (as hazard appears to be non-constant from the log-cumulative hazard plot) and Weibull for R-CHOP/R-CVP, going by AIC and BIC as the main criteria.

   For the R-mono comparison, the selected log-logistic curve did not seem to fit very well to the R-mono arm. No justification was provided for choosing log-logistic. In a scenario, the ERG explored the use of the generalised gamma, which fitted R-mono better (but was a worse fit to R^2).

g) For TTNLT, the main reason to select the log-normal curve was because the exponential curve would cause crossing of TTNLT and PFS arms. However, as the log-normal distribution did not fit the R^2 data very well, and the crossing of curves would not actually take place but was corrected for, the ERG considered the exponential curve equally suitable. The choice of TTNLT curve was however not too influential as there are no consequences for OS and utility scores in the company base-case were high throughout, making TTNLT mostly about the timing of the one-off subsequent treatment costs and a slight utility decrement. For the R-mono comparison, in the absence of any diagnostic plots, it was difficult to see whether the generalised gamma would be the optimal choice, but the ERG considered that given AIC and BIC for the various parametric survival curves, there

may not be one model type that would make a good fit to both arms and so the company base-case was left unchanged at this point.

h) When running the scenario where efficacy data from van Oers were used, the model returned error values. This was caused by a dot versus comma issue (possibly specific to the version of Excel used to run the model in) in the parameters for the parametric survival curves. In the ERG base-case, this error was fixed.

Figure 5.2: PFS curves for R^2 with in addition the KM + comparator hazard curve that was used in base-case

Source: adapted from company model[26] – KM + comparator curve added

Figure 5.3: PFS curves for R-CHOP/R-CVP (GenGamma used in company base-case, log-logistic in ERG base-case)

Source: adapted from company model[1] – changed line presentation styles

5.2.7 Adverse events

The main sources of evidence on treatment-related adverse events used for intervention and comparators are the AUGMENT[64] and RELEVANCE[66] trials, because of a lack of safety data from HMRN. RELEVANCE is a Phase III study comparing R^2 with R-chemotherapy (R-CHOP, R-CVP and R-Benda) for patients with previously untreated FL. AUGMENT was used in the base-case analysis for R^2, and RELEVANCE was used for R-CHOP and R-CVP with incidence adjusted for relative incidence of R^2 in AUGMENT compared to R^2 in RELEVANCE:

Comparator AE incidence = ($AE_{comparator}$ incidence in RELEVANCE/AE_{R2} incidence in RELEVANCE) x AE_{R2} incidence in AUGMENT.

So, the incidence of R-CHOP and R-CVP AEs were adjusted for any possible differences in R^2 AEs between AUGMENT and RELEVANCE.

Grade 3/4 AEs with incidence of greater than 2% in either treatment were considered. If any reported AEs for R-CHOP/R-CVP were >2% incidence, they were also included for R^2. Any AEs reported in AUGMENT that were used in the model, but were not reported for the comparator, were assumed 0% incidence for the comparator and not costed for.

In a scenario, AEs for the comparators were taken from van Oers et al. (2006)[39] which concerned a relapsed/refractory population. As van Oers et al. was a study on R-CHOP and no data on AEs in R-CVP were available, in this scenario the R-CHOP AE incidences were also applied to the R-CVP comparator.

Furthermore, AE incidence for maintenance treatment and autologous stem cell transplant (ASCT) were also considered. The incidence of AEs for rituximab maintenance were taken from van Oers et al. (2010)[67] and were neutropenia (11.5%) and infection (19.7%). In line with NG52 NHL guidelines,[68] the

only post ASCT AE for the model was febrile neutropenia, for 98.3% of patients undergoing ASCT as taken from Leger et al.[69]

ERG comment: The main concerns of the ERG relate to: a) the omission of AEs related to ASCT and subsequent R-mono therapy in the R^2 arm b) the RELEVANCE population being exclusively patients that were previously untreated.

a) For patients in the R-CHOP/R-CVP arm undergoing ASCT and R-mono as subsequent therapy, AEs related to these treatments were accounted for and costed in the model. Also, a small utility decrement was applied for these AEs. For the R^2 arm, even though these therapies were also observed here (be it to a lesser extent), related AEs were not accounted for. The ERG considered this to be inconsistent and corrected for it in the ERG base-case.

b) AE incidences in the company base-case were taken from the RELEVANCE trial[66] which concerned a population of previously untreated patients. Data from a relapsed/refractory population were only used in a scenario, with lower AE incidences than in the base-case. The ERG questions the applicability of RELEVANCE for the present STA. On the one hand, it may be the case that previously untreated patients have fewer side effects than a relapsed/refractory population, since they have not built up any intolerances. On the other hand, one would expect that a population receiving second-line treatment might be a special selection in the sense that those patients who experienced severe AEs in first-line will not be eligible for second-line. Either way, the ERG feels it is important to seriously consider the scenario provided by the company. Therefore, the ERG included it as one of their scenarios.

5.2.8 Health-related quality of life

The utility values were estimated for the health states PF, and PP off and on treatment using EQ-5D-3L data collected in AUGMENT. A covariate selection process was used to select the appropriate mixed effects utility model as input for the economic model. The final covariates included in the model were health state (PF versus PP), next anti-lymphoma treatment, treatment, baseline utility, previous rituximab exposure, refractory to last prior regimen and number of prior therapies. The R^2 arm had a utility increment of 0.011 compared with the R-mono arm for all health states. However, given that this difference was minimal and not statistically significant, the company used the same utility values based on R^2 in the model.

Health-related quality of life data identified in the review

According to the CS, the SLR identified a total of 38 studies from 53 publications, including 12 HTAs and one observational study. Out of these, the company considered the utility values of the studies of Wild et al., Pereira et al. and TA472 most relevant.[20, 70, 71]

Health state utility values

The utility values resulting from the mixed effects model were used to inform the health states in the model for all treatments, and utility values from the study of Wild et al.[70] were tested in a scenario analysis. However, the disease characteristics that were used to derive utility values from the mixed effects model were population-dependent, and therefore, the utility values for R^2 versus R-CHOP/R-CVP and R^2 versus R-mono are slightly different. A summary of all utility values used in the model is provided in Table 5.4. The company stated in the CS that the mean utility value for the PF state was generally consistent with those reported in the three studies selected from the SLR, with the exception of the lower PF utility value of Pereira et al.[71] The mean utility values for post-progression were higher based on AUGMENT trial data compared with the other studies.

Table 5.4: Health state utility values

State	Utility value R^2 versus R-CHOP/CVP	Utility value R^2 versus R-mono	Reference	Justification
PF	0.863	0.847	Section B.3.4, page 177 and 182 of the CS.	EQ-5D values derived from a relevant patient population and model specific health states.
PP (off treatment)	0.837	0.821		
PP (on treatment)	0.808	0.792		
Source: based on Table 47 of the CS				

Adverse event related disutility values

Utility decrements for grade 3 and 4 AEs were applied in the model for the expected duration of each AE, based on literature and previous appraisals. See Table 5.5 for details on the AE utility decrements, durations and sources.

Table 5.5: Adverse event related disutility values

Adverse event	Disutility value	Duration (days)	Source for disutility	Source for duration
Neutropenia	0.090	15.09	Nafees et al. (2008)[72]	TA306[73]
Leukopenia	0.119	13.96	TA513 (assumed to be the same as anaemia)[23]	TA306[73]
Anaemia	0.119	16.07	Swinburn et al. (2010)[74]	TA306[73]
Pneumonia	0.200	14.00	Beusterien et al. (2010)[75]	TA306[73]
Lymphocyte count decreased	0.100	34.00	Stein et al. (2018)[76]	Assumed maximum of all Grade 3/4 AEs
Lymphopenia	0.100	34.00	Stein et al. (2018)[76]	TA306[73]
Febrile neutropenia	0.150	7.14	Lloyd et al. (2006)	TA306[73]
White blood cell count decreased	0.100	34.00	Stein et al. (2018)[76]	Assumed maximum of all Grade 3/4 AEs
Diarrhoea	0.048	34.00	Nafees et al. (2008)[72]	Assumed maximum of all Grade 3/4 AEs
Thrombocytopenia	0.108	23.23	Tolley et al. (2013)[77]	TA306[73]
Hypokalaemia	0.124	34.00	TA423[78]	Assumed maximum of all Grade 3/4 AEs

Adverse event	Disutility value	Duration (days)	Source for disutility	Source for duration
Pulmonary embolism	0.124	34.00	TA423[78]	Assumed maximum of all Grade 3/4 AEs
Infusion-related reaction	0.195	34.00	Tolley et al. (2013)[77]	Assumed maximum of all Grade 3/4 AEs
Nausea and emesis	0.048	6.00	Nafees et al. (2008)[72]	TA306[73]
Allergic reaction	0.098	34.00	Hannouf et al. (2012)[79]	Assumed maximum of all Grade 3/4 AEs
Hypotension	0.057	8.00	Hannouf et al. (2012)[79]	TA306[73]
Fatigue	0.073	31.50	Nafees et al. (2008)[72]	TA306[73]
Alopecia	0.045	34.00	Nafees et al. (2008)[72]	Assumed maximum of all Grade 3/4 AEs
Infection	0.195	34.00	Tolley et al. (2013)[77]	Assumed maximum of all Grade 3/4 AEs
Sepsis	0.267	34.00	Hannouf et al. (2012)[79]	Assumed maximum of all grade ¾ AEs
Abdominal pain	0.069	17.00	Doyle et al. (2008)[80]	TA306[73]
Acute kidney injury	0.270	29.75	TA306[73]	TA306[73]
Source: Based on Table 44 of the CS.				

ERG comment: The main concerns of the ERG relate to: a) the high utility values for the PF and PP off treatment and the PP on treatment health states; and b) the modest utility decrement for progressed disease;

a) Utility values for the PF (0.863 for R^2 versus R-CHOP/R-CVP and 0.847 for R^2 versus R-mono) and PP (off treatment 0.837 for R^2 versus R-CHOP/R-CVP and 0.821 for R^2 versus R-mono, on treatment 0.808 for R^2 versus R-CHOP/R-CVP and 0.792 for R^2 versus R-mono) health states were higher than the utility reported for the general population (0.80 for age category 55-64).[81] Utility scores higher than in the general population seem quite unlikely in patients with treated FL or MZL. In addition, these utility values were also higher than reported in the literature for this population.[70, 71] Also, the company decided to go with the slightly higher utilities from the R^2 arm in AUGMENT, even though there was not a significant difference between R^2 and R-mono. The ERG capped the utility values in its base-case to general population norms, which had a low impact but slightly increased the ICER.

b) The utility difference between the PF health state and the PP off treatment and PP on treatment health states were -0.026 and -0.056 respectively in the R^2 versus R-CHOP/R-CVP comparison and

respectively -0.026 and -0.055 in the R^2 versus R-mono comparison. This seems modest given the difference in utility value between these health states reported in the literature, which show differences up to -0.27.[71] The ERG judges that a larger utility difference between PF and PP health states would be more plausible, and explored this in a scenario analysis using utility values of Wild et al. (0.62) and Pereira et al. (0.45) for both PP health states. For R^2 versus R-CHOP and R/CVP, this substantially increased the ICER, while for R^2 versus R-mono the ICER decreased.

5.2.9 Resources and costs

The cost categories included in the model were costs associated with treatment (drug acquisition costs including subsequent therapies, drug administration costs including subsequent therapies, costs associated with treatment-related AEs), disease monitoring costs and costs associated with end of life care.

Unit prices were based on the National Health Service (NHS) reference prices,[82] Personal Social Services Research Unit (PSSRU),[83] the Monthly Index of Medical Specialities (MIMS)[84-90] and the Electronic Market Information Tool (eMIT)[91].

Resource use and costs data identified in the review

According to Appendix I of the CS[1], the SLR identified 17 studies of which 14 reported UK relevant resource use and cost information. The CS did not state which of these studies the company considered to be consistent with the NICE reference case and appropriate for the economic model.

Drug acquisition costs (with PAS)

For lenalidomide, dosing data had been taken directly from AUGMENT (non-R-refractory population) to align the drug costs with the efficacy data because according to the company, dose reductions for lenalidomide can occur. To capture the impact of treatment reductions or missed treatment cycles over time on costs, the observed number of patients on each dosage at every cycle was combined with the unit drug costs to calculate a weighted cost per cycle. This cost was then multiplied by the proportion of patients eligible for treatment who receive treatment in that cycle (based on ToT KM curves and the mean treatment cycle length). To align with the costing method applied for lenalidomide, the same method was applied to calculate rituximab costs for the R^2 arm. The use of mean relative dose intensities (RDIs) were explored in scenario analyses, using values of [redacted] and [redacted] for rituximab and lenalidomide in the R^2 arm of AUGMENT, respectively (Table 5.6).

The proportion of patients eligible for treatment who receive treatment in each arm in the rituximab monotherapy arm of AUGMENT was applied to all comparators in the model, in order to similarly align the costing of the comparators to the study dosing methods described above for R^2. A mean dose intensity value of 87.5% was assumed in scenario analyses across all individual chemotherapies within the R-chemotherapy comparator regimens. No dose intensity value was applied to rituximab within R-chemotherapy combinations or R-maintenance, because dose reductions were not recommended for rituximab. For BSA dependent treatments, the company applied the method of moments technique to IPD from AUGMENT to calculate the average number of vials that would be required to satisfy one administration of treatment. Other methods, such as dose banding and using the minimum cost per mg for each treatment (no wastage), were explored in scenario analyses.

Patients in the R^2 received allopurinol (in the first treatment cycle only) and filgrastim as concomitant treatments. Rituximab maintenance was given every three months up to two years or until disease progression to patients who responded to R-chemotherapy induction treatment.

Table 5.6: Treatment acquisition costs

Treatment	Size	Cost per pack	Source
Lenalidomide (with PAS)	21 x 2.5 mg tablets	[redacted]	MIMS (Revlimid)[85]
	21 x 5 mg tablets	[redacted]	
	21 x 10 mg tablets	[redacted]	
	21 x 15 mg tablets	[redacted]	
	21 x 20 mg tablets	[redacted]	
Rituximab	2 x 100 mg vials	£349.25	MIMS (MabThera)[84]
	1 x 500 mg vial	£873.15	
	1 x 1,400 mg (SC)	£1,344.65	
	2 x 100 mg vials	£314.33	MIMS (Rixathon)[89]
	1 x 500 mg vial	£785.84	
Cyclophosphamide	1 x 1,000 mg vial	£13.47	eMIT[91]
	1 x 2,000 mg vial	£27.50	
	1 x 500 mg vial	£8.31	
Doxorubicin	1 x 10 mg vial	£4.48	
	1 x 200 mg vial	£15.59	
	1 x 50 mg vial	£17.78	
Vincristine	5 x 1 mg vials	£11.59	
	5 x 2 mg vials	£17.82	
	5 x 5 mg vials	£99.00	
Prednisolone	28 x 1 mg tablets	£0.17	
	28 x 2.5 mg tablets	£0.59	
	30 x 20 mg tablets	£4.17	
	56 x 25 mg tablets	£20.25	
	28 x 5 mg tablets	£0.27	

Source: based on Table 49 of the CS.

eMIT = electronic market information tool; MIMS = Monthly Index of Medical Specialities; PAS = patient access scheme; SC = subcutaneous

Administration costs

Drug administration costs were based on NHS reference costs tariffs, pharmacy costs for the preparation of the infusion, and NHS transport costs.[82] For rituximab combination chemotherapies, a cost of £374.52 was applied at first administration of each cycle, followed by a cost of £312.34 for subsequent administrations per cycle. For simpler chemotherapies such as in the R^2 arm, first administration per cycle cost £309.22 and £312.34 for subsequent administrations per cycle. For all infusion treatments, pharmacy costs were applied assuming a 15-minute infusion preparation time based on TA243[21] and £48 per hour for hospital-based scientific and professional staff from PSSRU costs.[83] NHS transport costs were assumed in 30% of patients and were applied to all administrations in the model.

Treatment-specific monitoring

Costs of a full blood count were added to each treatment cycle for lenalidomide per visit to monitor the dose-limiting toxicities of neutropenia and thrombocytopenia.

Health state costs

Table 5.7 presents the costs that are included in the economic model per health state.

Table 5.7: Health state related costs

Health state	Costs	Cost components considered	Reference
PF (on treatment)	Drug acquisition	**R²:** Cycle 1: £[redacted] Cycles 2-5: £[redacted] Cycles 6-12: £[redacted] **R-CHOP:** £1,216 per cycle **R-CVP:** £1,200 per cycle **R-mono:** Cycle 1: £4,680 Cycles 2-5: £1,170	Table 48, page 187 of the CS
	Drug administration	**R²:** Cycle 1: £1,348 Cycles 2-5: £335 **R-CHOP/R-CVP:** £400 per cycle **R-mono:** Cycle 1: £1,348 Cycles 2-5: £335	Table 48, page 187 of the CS
	Maintenance/ASCT	R-maintenance: £1,345 (SC), £1,170 (IV)	Table 48, page 187 of the CS
		ASCT: £35,558	Table 59, page 202 of the CS
	Disease monitoring	£254,95 per month	Table 57, page 200
	Adverse events	£1,832 (**R² non-R-refractory**) £3,604 (**R-CHOP**) £2,754 (**R-CVP**) £462 (**R-mono**) £1,773 (R^2 R-refractory) £370 (R-maintenance) £6,336 (ASCT)	Table 61, page 204 of the CS
PF (off treatment)	Disease monitoring	£83.09 per month	Table 57, page 200 of the CS

Health state	Costs	Cost components considered	Reference
PP (on treatment)	Disease monitoring	£232.17 per month	Table 57, page 200 of the CS
	Subsequent treatments	£5,195 (**R^2**) £8,371 (**R-CHOP/R-CVP**)	Table 62, page 206 of the CS
PP (off treatment)	Disease monitoring	£58.04 per month	Table 57, page 200 of the CS
Death	Terminal care	£6,362	Page 206 of the CS

Source: Based on Table 56 of the CS
CHOP = cyclophosphamide, doxorubicin hydrochloride, vincristine and prednisolone; CVP = cyclo-phosphamide, vincristine and prednisolone; R = rituximab; R^2 = lenalidomide plus rituximab.

Disease monitoring

Disease monitoring resource use costs were assumed to be similar to those in previous FL submissions[20, 21, 23] and were split by health state. Patients were assumed to have monthly haematologist visits and diagnostic tests with a CT scan every six months in the induction phase of the PF health state. In the maintenance phase of the PF health state, follow-up visits (based on ESMO guidelines) were reduced to one visit every three months and one annual CT scan, and in the post-maintenance phase to one visit every four months without CT scan. In the PP state, higher visit and diagnostic test frequencies were assumed (both monthly). Although resource use information for MZL was limited, similar tests and frequencies were suggested in the MZL ESMO guidelines[29] and therefore disease monitoring costs were assumed to be identical to FL.

Stem cell transplant (pre-progression)

Patients that were fit and young enough and who relapse early but who are not refractory to induction therapy were considered for consolidation with ASCT. In the economic model, the company applied ASCT to [redacted]% of patients in R-CHOP. For R-CVP, 0% ASCT was applied as R-CVP was considered unlikely to be used as an induction regimen prior to ASCT. For R^2, as clinicians suggested that it was unlikely that ASCT would be offered post R^2 in clinical practice and ASCT was also not offered to patients after R^2 within the AUGMENT protocol, 0% ASCT was applied as well. The cost of ASCT was based on NHL guidance uplifted to 2018 costs and included £35,558.15.[25, 83] The NHS reference cost (£18,520.20) for ASCT was used in a scenario analysis.

Adverse event related costs and costs of terminal care

The frequency of grade 3-4 AEs that occurred in ≥2% of patients was applied to the incidence rate for each treatment to obtain a one-off upfront cost to each treatment arm in the model.

Furthermore, to reflect the costs of terminal care, a one-off cost of £6,361.77 was applied in the model when a patient died. This cost was based on the average cost derived from the Round et al. (2015) modelling study,[92] which estimated the cost of cancer care during the final phases of life.

Total AE costs per treatment are shown in Table 5.8.

Table 5.8: Total AE costs per treatment

Treatment	Total costs
Non-R-refractory	
R^2	£1,831.71
R-CHOP	£3,604.13
R-CVP	£2,753.56
R-mono	£462,41
R-refractory	
R^2	£1,773.94
Post-induction	
R-maintenance	£369.95
ASCT	£6,400.93

Source: based on Table 61 of the CS.
AE = adverse event; ASCT = autologous stem-cell transplant; CHOP = cyclophosphamide, doxorubicin hydro-chloride, vincristine and prednisolone; CVP = cyclophosphamide, vincristine and prednisolone; R = rituximab; R^2 = lenalidomide plus rituximab.

Costs of subsequent treatments

Subsequent treatments were included in the model as an average one-off cost to patients entering the PP (on treatment) health state, derived using TTNLT data. Costs for patients in the R^2 arm were derived from subsequent treatments from AUGMENT. The total subsequent treatment data from the pooled R-chemotherapies in the HMRN database were used for R-CHOP and R-CVP. The company also conducted a scenario analysis in which the costs were equalised by applying the subsequent treatment costs of the comparator arm to R^2. The mean duration of subsequent treatments was based on HMRN data, with AUGMENT mean durations used in a scenario analysis.

ERG comment: The main concerns of the ERG relate to: a) subsequent treatments that were included as a one-off cost and were therefore potentially underestimated; b) the source used for the proportion of patients who receive subsequent treatment after R-CHOP/R-CVP to determine subsequent treatment costs; and c) the omission of data observed in AUGMENT to inform pre-progression ASCT in the R^2 arm.

a) Subsequent treatments were included in the model as a one-off cost to those patients entering the PP on treatment health state. The company costed for observed incidences of subsequent treatments from the data sources, which for R2 had a much shorter follow-up than for R-CHOP/R-CVP. The ERG is concerned that because of the limited follow-up in AUGMENT as compared to HMRN, this assumption does not reflect clinical practice and subsequent treatment costs for R^2 in the economic model are therefore likely to be underestimated. Although subsequent treatment duration in the model lasts no longer than a maximum of 130.3 days, patients in the PP on treatment health state remain in this health state until they die, and the relatively high age-adjusted utility value corresponding to this health state is assumed over this whole time span. The ERG is concerned about the fact that subsequent treatment costs, in contrast to the utilities, are not counted over the remaining time that patients stay in the PP on treatment health state. As patients in the R^2 arm

remain in the PP on treatment health state for a longer time on average, applying subsequent treatment costs as one-off possibly favoured R^2.

b) To calculate subsequent treatment costs, the company based the proportion of patients receiving subsequent treatment after R-CHOP/R-CVP on the total subsequent treatment data from the pooled R-chemotherapies from HMRN because of its larger sample size (n=129) compared to the HMRN R-CHOP/R-CVP cohort (n=67). However, the ERG judges that, in line with the treatment effectiveness, the R-CHOP/R-CVP cohort should be used to calculate subsequent treatment costs and applied this to the ERG base-case for the comparison with R-CHOP and R-CVP. This resulted in slightly lower subsequent treatment costs for R-CHOP and R-CVP and a slightly higher ICER, although the impact was small. In addition, the ERG also explored the impact of equal subsequent treatment costs between R^2 and R-CHOP, R-CVP and R-mono, which resulted in a large increase of the ICER.

c) The company assumed the percentage of post-induction (but pre-progression) ASCTs in R^2 to be zero, because it was not protocolised in AUGMENT and clinicians considered it unlikely that patients would receive ASCT post R^2. The ERG was unable to find any report of actual incidence of ASCT performed post R^2 in AUGMENT, but would have liked to see a scenario using observed frequencies, as clinical practice may sometimes contrast with protocols and clinical opinion. If the observed frequency was non-zero, this would increase the ICER for R^2 compared to R-CHOP.

5.2.10 Cost effectiveness results

R^2 versus R-CHOP and R-CVP

In the deterministic base-case analysis, total LYs and QALYs gained were larger for R^2 than for R-CHOP and R-CVP. Incremental QALYs ([redacted]) were mainly driven by QALY gains in the PP (off treatment) health state. Total costs were also higher for R^2 than for R-CHOP and R-CVP. Incremental costs ([redacted] and [redacted] respectively) mainly resulted from higher drug acquisition (induction) costs. The deterministic incremental cost effectiveness ratio (ICER) amounted to £11,471 per QALY gained for R^2 versus R-CHOP and £16,814 for QALY gained for R^2 versus R-CVP (see Table 5.9).

R^2 versus R-mono (added by the company after the clarification phase)

In the deterministic base-case analysis, total LYs and QALYs gained were larger for R^2 than for R-mono. Incremental QALYs ([redacted]) were mainly driven by QALY gains in the PF health state. Total costs were also higher for R^2 than for R-mono. Incremental costs ([redacted]) mainly resulted from higher drug acquisition (induction) costs. The deterministic cost effectiveness ratio (ICER) amounted to £22,580 per QALY gained (see Table 5.9).

Table 5.9: Company's deterministic base-case results

	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER (£/QALY)
R^2 versus R-CHOP							
R^2	£[redacted]	[redacted]	[redacted]	--	--	--	--
R-CHOP	£[redacted]	[redacted]	[redacted]	£[redacted]	[redacted]	[redacted]	£11,471
R^2 versus R-CVP							
R^2	£[redacted]	[redacted]	[redacted]				
R-CVP	£[redacted]	[redacted]	[redacted]	£[redacted]	[redacted]	[redacted]	£16,814

	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER (£/QALY)
R^2 versus R-mono							
R^2	[redacted]	[redacted]	[redacted]	--	--	--	--
R-mono	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	£22,580
Source: Based on Table 64 of the CS ICER = incremental cost effectiveness ratio; LYG = life years gained; QALY = quality adjusted life year.							

5.2.11 Sensitivity analyses

The company performed a probabilistic sensitivity analysis (PSA) and deterministic sensitivity analysis (DSA) to show the uncertainty surrounding the base-case results.

R^2 versus R-CHOP and R-CVP

Compared with the deterministic results, the PSA with 1,000 iterations showed lower incremental QALYs and costs for both R-CHOP and R-CVP, which resulted in increased ICERs of £13,443 (versus R-CHOP) and £20,896 (versus R-CVP) (see Table 5.10). The cost effectiveness acceptability curve in the economic model showed that R^2 had an 82% (versus R-CHOP) and 72% (versus R-CVP) probability of being cost effective at a willingness-to-pay (WTP) threshold of £30,000.

The company performed DSAs by varying key model parameters between their upper and lower limits of the confidence intervals. For R^2 versus R-CHOP, the ICER was most sensitive to the cost of ASCT, the total subsequent treatment costs for R-CHOP and the proportion of patients who receive SCT. For R^2 versus R-CVP, the ICER was most sensitive to the total subsequent treatment costs for R-CVP (including ASCT costs) and resource use costs. For both comparisons, in none of the DSAs the ICER exceeded the WTP threshold of £30,000.

R^2 versus R-mono (added by the company after the clarification phase)

For R^2 versus R-mono, the company only provided basic deterministic results and the PSA and DSA were performed by the ERG. Compared with the deterministic results, the PSA with 1,000 iterations showed lower incremental QALYs and costs, which resulted in an increased ICER of £26,116) (see Table 5.10). The cost effectiveness acceptability curve in the economic model showed that R^2 had a 69% probability of being cost effective at a willingness to pay (WTP) threshold of £30,000.

The company performed DSAs by varying key model parameters between their upper and lower limits of the confidence intervals. The ICER was most sensitive to the total subsequent treatment costs for R^2 and R-mono and the frequency of haematologist visits post progression. In none of the DSAs the ICER exceeded the WTP threshold of £30,000.

Table 5.10: Company's base-case results (probabilistic, 1,000 iterations)

	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	ICER (£/QALY)
R^2 versus R-CHOP					
R^2	£[redacted]	[redacted]	--	--	--
R-CHOP	£[redacted]	[redacted]	£[redacted]	[redacted]	£13,443
R^2 versus R-CVP					
R^2	£[redacted]	[redacted]			
R-CVP	£[redacted]	[redacted]	£[redacted]	[redacted]	£20,896

	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	ICER (£/QALY)
R^2 versus R-mono					
R^2	£[redacted]	[redacted]	--	--	--
R-mono	£[redacted]	[redacted]	£[redacted]	[redacted]	£26,116
Source: Based on Table 64 of the CS. ICER = incremental cost effectiveness ratio; QALY = quality adjusted life year.					

Scenario analyses

The company conducted several scenario analyses. The results for R^2 versus R-CHOP showed ICERs ranging between £4,398 and £14,891 per QALY gained, excluding the scenarios assessing different time horizons. The three most influential scenarios that decreased the ICER were using an exponential distribution for R^2 ToT (£4,398), a 0.0% discount rate for QALYs (£8,174) and using a log-normal distribution for R^2 ToT (£8,312).

The results for R^2 versus R-CVP showed ICERs ranging between £9,731 and £20,636 per QALY gained, excluding the scenarios assessing different time horizons. The three most influential scenarios that increased the ICER were a 6.0% discount rate for QALYs (20,636), applying the comparator hazard to R^2 arms after three years (£20,471) and using a Gompertz distribution for R^2 PFS (£20,413). The three most influential scenarios that decreased the ICER were using exponential (£9,731) and log-normal (£13,650) distributions for R^2 ToT and a 0.0% discount rate for QALYs (11,976). For R^2 versus R-mono, the ICERs ranged between £12,125 and £43,814, excluding the scenarios assessing different time horizons. The three most influential scenarios that increased the ICER were applying the comparator hazard to R^2 arms after three years (£43,814), using a 6.0% discount rate for QALYs (£27,613) and applying the same subsequent treatment costs for R^2 and R-mono (£24,951). The three most influential scenarios that decreased the ICER were applying the comparator hazard to R^2 arms after 10 years (£12,125), using an exponential distribution for R^2 ToT (£13,845) and using a 0.0% discount rate for QALYs (£16,391).

ERG comment: The main comments of the ERG relate to: a) the inability to perform a fully incremental analysis in the model; b) unstable PSA results; and c) the additional scenario analysis for the FL-only population.

a) The PSA did not enable simultaneous calculation of outcomes for more than two comparators and representation of multiple comparators in the cost effectiveness acceptability curve (CEAC). Therefore, the ERG created three separate model files. Furthermore, compared with the company's deterministic base-case results, probabilistic incremental QALYs are lower, likely caused by non-linearity of the model.

b) The ERG twice performed a PSA with 10,000 iterations to test its stability, but increasing the number of iterations did not stabilise the results.

c) An additional scenario analysis for the FL-only population was provided by the company in response to clarification. The FL-only scenario resulted in increased deterministic ICERs of £15,909 and £23,746 for the R-CHOP and R-CVP comparisons, respectively, making it the most influential scenario. For the R-mono comparison, using FL-only data lowered the ICER to £20,310.

Therefore, given that the pooling of FL and MZL population appeared to have a substantial impact, the ERG included the FL-only scenario in their exploratory analyses.

5.2.12 Model validation and face validity check

Face validity

The model structure and its appropriateness to reflect the clinical pathway, notably the decision to split the progressed disease health state up into on- and off-treatment, were validated in an advisory board consisting of six clinicians and two UK economic experts. These further validated the extrapolation of survival beyond the trial period, the indirect treatment comparison, the use of clinical validity of utilities derived from AUGMENT versus those in the literature and subsequent treatment usage.

Internal validity

Distributions to estimate PFS, OS and TTNLT were chosen such that no implausible curve crossing occurred. A health economist that was not involved in model development reviewed the model for coding errors, inconsistencies and input plausibility. Several extreme value checks were also performed and sub-modules of the model were tested.

Cross validity

No cross validity checking of the model was reported by the company, although the company did state that the chosen modelling approach of partitioned survival analysis with the health states of PF, PP and dead was in line with "the majority of economic evaluations found in the SLR".[26] However, the company then diverted from this path by adding additional health states (splitting progression by whether patients were on or off treatment given that TTNLT was considered a better endpoint than PFS).

External validity

Model predictions for PFS, OS, TTNLT were compared with the respective KM data from AUGMENT and found mostly in line, with the notable exception of 1-year PFS for R-CHOP/R-CVP that was under-estimated in the model compared to the observed data. According to the company, from two years onwards model predictions were more aligned with observations for this outcome. Comparisons with other trials were not made because no other datasets were available.

Predictive validity

No predictive validity checking was reported by the company.

ERG comment: The main concerns of the ERG relate to: a) limited information available on the company's validation efforts based on the CS and b) concerns regarding external validity.

a) The company provided limited information on its validation efforts. In response to the clarification letter, however, the company provided the meeting report of the advisory board[26] and the filled in Assessment of the Validation Status of Health Economic Decision Models (AdViSHE) tool.[93] The latter shed more light on especially the internal validation of the company's model, which was performed to a good standard. The advisory board meeting report supported some model approaches and assumptions, but not all: for instance, the model structure including the on- and off-treatment division was not corroborated, and neither was the choice of distributions for R-CHOP/R-CVP OS and PFS.

b) External validation exercised by the company found that R-CHOP/R-CVP PFS, OS and TTNLT at one year were under-estimated. Whilst these extrapolations stabilised from two years onwards to be more aligned with the observed data, this under-estimation may still have an impact on cost effectiveness estimates, as explored in the treatment effectiveness section. Furthermore, it is not clear whether these extrapolations have been validated by experts as the expert meeting minutes only contained a statement regarding (the comparison with) R-mono.

5.3 Exploratory and sensitivity analyses undertaken by the ERG

Table 5.11 summarises the main issues highlighted by the ERG in Section 5.2 of this report, indicates the expected direction of bias introduced by these issues and whether these are examined in any analyses/incorporated in the ERG base-case.

Table 5.11: Main ERG critique of company's submitted economic evaluation

Issue	Likely direction of bias introduced in ICER[a]	ERG analyses	Addressed in company analysis?
Model structure (section 5.2.2)			
Partitioned survival analysis, no alternative results from state transition model provided for comparison	+/-	No	Requested but not provided
Population, interventions and comparators, perspective and time horizon (sections 5.2.3-5.2.5)			
O-Benda not a relevant comparator, while R-mono left out	NA	Base-case	R-mono analysis provided by company upon request
MZL and FL populations were pooled throughout the analyses, as assumed to be similar by the company	+/-	Scenario	Scenario provided upon request
Treatment effectiveness and extrapolation (section 5.2.6)			
Indirect comparison seems to inflate R^2 efficacy and lower costs relative to R^2 in direct comparison based on AUGMENT	+	No	No
Substantial uncertainty concerning extrapolation of PFS curves. Company base-case not based on best fit, nor solid other justification	+	Base-case (MJ), scenarios	Scenarios
Curves for OS extrapolation do not provide best fit, choice is not sufficiently justified	+/-	Base-case (MJ), scenarios	Scenarios
Curves for TTNLT extrapolation do not provide best fit, choice is not sufficiently justified	+/-	Base-case (MJ), scenarios	Scenarios
Cut-off for treatment effectiveness at 5 years not supported by evidence	+/-	Scenarios	Scenarios
Adverse events (section 5.2.7)			
Incidence for adverse events in R-CHOP and R-CVP taken from published source on a previously untreated population	+/-	Scenario	Scenario
AEs (costs and utility decrements) related to subsequent treatments were omitted for R^2	+	Base-case (FV)	No

Issue	Likely direction of bias introduced in ICER[a]	ERG analyses	Addressed in company analysis?
Health-related quality of life (section 5.2.8)			
Utility scores for all health states are likely high	+	Base-case (FV)	Yes, scenarios allow for alternative values
Utility decrement post progression low	+	Scenarios	
Resources and costs (section 5.2.9)			
One-off costs for subsequent treatment likely underestimates R^2 costs	+	Scenario	Scenario using same subsequent treatment costs
Incidences of subsequent treatments for R-CHOP and R-CVP were taken from the mixed R-chemo group of HMRN, which likely is an overestimate	+	Base-case (FV)	No
Consolidation ASCT in R^2 arm assumed zero, data on observed number of ASCTs was not provided in CS	+	No	No
Cost effectiveness analyses (sections 5.2.10 and 5.2.11)			
Discrepancy between probabilistic and deterministic results	+/-	No	No
PSA does not allow for full incremental analysis	+/-	No	No

Footnotes: [a] Likely conservative assumptions (of the intervention versus all comparators) are indicated by '-'; while '+/-' indicates that the bias introduced by the issue is unclear to the ERG and '+' indicates that the ERG believes this issue likely induces bias in favour of the intervention versus at least one comparator.
ERG = Evidence Review Group; FE = Fixing errors; FV = fixing violations; ICER = incremental cost effectiveness ratio; MJ = matters of judgement; NA = not applicable.

Based on all considerations in Section 5.2 of this report (summarised in Table 5.11), the ERG defined a new base-case. This base-case included multiple adjustments to the original base-case presented in the previous sections. These adjustments made by the ERG form the ERG base-case and were subdivided into three categories (derived from Kaltenthaler 2016[94])

- Fixing errors (correcting the model where the company's submitted model was unequivocally wrong)
- Fixing violations (correcting the model where the ERG considered that the NICE reference case, scope or best practice had not been adhered to)
- Matters of judgement (amending the model where the ERG considers that reasonable alternative assumptions are preferred).

The adjustments apply to the R-CHOP and R-CVP comparisons. For the R-mono comparison, these adjustments may be different or do not apply. In the list below, when nothing is mentioned on R-mono, this implies that this particular adjustment was similarly applied to the R-mono comparison.

Fixing errors

1. Error cells when using 'van Oers' as input for R-CHOP efficacy (section 5.2.6).
   The ERG replaced dots by commas in the van Oers parameters for curves.

Fixing violations

2. AEs related to subsequent treatments not accounted for in R^2 arm (section 5.2.7).
   The ERG included costs and utility decrement for AEs related to ASCT and rituximab subsequent treatment in R^2 arm like in the comparator arm.
3. Subsequent treatment rates for R-CHOP/R-CVP taken from mixed R-chemo population (section 5.2.9). The ERG used pooled R-CVP/R-CHOP subsequent treatment rates instead of R-chemo. (Not applicable in the R-mono comparison)
4. Utilities in all health states were higher than or comparable to general population levels (section 5.2.8). The ERG capped utilities at the general population level

Matters of judgment

5. Weibull distributions for OS do not provide the best fit, reasons for selecting unclear. (section 5.2.6). The ERG used the exponential distribution to extrapolate OS in both arms
6. KM+comparator hazard approach likely overestimates PFS in R^2 (section 5.2.6).
   The ERG used log-logistic for PFS in R^2 and Weibull for PFS in the comparator (not applied to R-mono comparison)
7. Lognormal curves for extrapolating TTNLT appear to be suboptimal (section 5.2.6).
   The ERG used log-logistic for TTNLT both arms (not applied to R-mono comparison)

Table 6.1 shows the combined effect of all abovementioned adjustments simultaneously, resulting in the (deterministic) ERG base-case. The 'fixing error' adjustments were combined and the other ERG analyses were performed also incorporating these 'fixing error' adjustments given the ERG considered that the 'fixing error' adjustments corrected unequivocally wrong issues.

5.3.1 ERG base-case results

The results of the deterministic ERG base-case are shown in Tables 5.12 and 5.13. The fully incremental analysis could only be performed for R^2, R-CHOP and R-CVP since when compared to R-mono, the effectiveness and costs of R^2 would be different. The ERG wishes to emphasise that all ERG analyses (except those for the R-mono comparison) are conditional upon the MAIC results for which uncertainty could not be quantified or incorporated in the economic model. For R-CHOP, deterministic incremental

costs were [redacted] and incremental QALYs [redacted] which resulted in an ICER of £15,505. Main drivers for the increased ICER compared to company base-case were the alternative OS and PFS curves, and to a lesser extent the use of only R-CHOP/R-CVP data for subsequent treatment rates, instead of pooled R-chemo data in the company base-case. For R-CVP, incremental costs were [redacted] and incremental QALYs were [redacted] (identical to the R-CHOP comparison), which resulted in an ICER of £21,759 which was driven by the same factors as in the R-CHOP comparison. Finally, for the R-mono deterministic comparison, incremental costs were [redacted] and incremental QALYs were [redacted] with a resulting ICER of £27,372. Main drivers were the cap of utilities to the level of the general population, and the use of alternative OS curves.

The fully incremental analysis showed R-CHOP to be strictly dominated, and the relevant comparison would be R^2 versus R-CVP. R-CHOP and R-CVP serve different populations however, and the ERG has already commented on the fact that pooling of R-CHOP and R-CVP may not be justified. If the assumption of equality does not hold, a fully incremental analysis based on a zero difference in QALYs between R-CHOP and R-CVP may not be indicated.

Compared with the deterministic base-case results, the ERG PSA with 1,000 iterations resulted in lower incremental costs but also in lower incremental QALYs, with consistently lower ICERs as a result, for all comparisons (see Table 5.14). For the R-CHOP comparison, the difference was quite modest with a probabilistic ICER of £15,818, but for R-CVP and R-mono it was more pronounced (probabilistic ICERs of £23,367 and £29,010, respectively). However, in the company base-case the differences between deterministic and probabilistic analyses were even larger, more than £4,000 in the R-CVP comparison for instance, and the cost effectiveness planes of the company base-case showed a number of extreme outliers concerning QALYs which were not observed in the ERG analyses (see Figure 5.4). This would imply the possibility that the QALY outliers in the company base-case may have been caused by the chosen distributions for extrapolating.

The cost effectiveness acceptability curves showed that compared to R-CHOP, R^2 approximately had an 83% and 90% probability of being cost effective at willingness-to-pay (WTP) thresholds of £30,000 and £50,000, respectively (See Figure 5.5). These percentages were lower for the R-CVP comparison; 68% and 84% (See Figure 5.6). For R-mono they were 54% and 77% (See Figure 5.7).

Table 5.12: ERG pairwise deterministic base-case results

Technologies	Total costs	Total QALYs	Incremental costs	Incremental QALYs	ICER (£/QALY)
Deterministic ERG base-case for R^2 versus R-CHOP					
R^2	[redacted]	[redacted]	[redacted]	[redacted]	£15,505
R-CHOP	[redacted]	[redacted]			
Deterministic ERG base-case for R^2 versus R-CVP					
R^2	[redacted]	[redacted]	[redacted]	[redacted]	£21,759
R-CVP	[redacted]	[redacted]			
Deterministic ERG base-case for R^2 versus R-mono					
R^2	[redacted]	[redacted]	[redacted]	[redacted]	£27,372
R-mono	[redacted]	[redacted]			

ERG = Evidence Review Group = ICER = incremental cost effectiveness ratio; QALY = quality-adjusted life year

Table 5.13: ERG fully incremental and pairwise deterministic base-case results for R2, R-CHOP and R-CVP (ICER compared to next relevant alternative)

Technologies	Total costs	Total QALYs	Incremental costs	Incremental QALYs	ICER (£/QALY)
R-CVP	[redacted]	[redacted]			
R-CHOP	[redacted]	[redacted]	[redacted]	[redacted]	Dominated
R^2	[redacted]	[redacted]	[redacted]	[redacted]	£21,759
ERG = Evidence Review Group = ICER = incremental cost effectiveness ratio; QALY = quality-adjusted life year					

Table 5.14: ERG probabilistic base-case results

Technologies	Total costs	Total QALYs	Incremental costs	Incremental QALYs	ICER (£/QALY)
Probabilistic ERG base-case for R^2 versus R-CHOP					
R^2	[redacted]	[redacted]	[redacted]	[redacted]	£15,818
R-CHOP	[redacted]	[redacted]			
Probabilistic ERG base-case for R^2 versus R-CVP					
R^2	[redacted]	[redacted]	[redacted]	[redacted]	£23,367
R-CVP	[redacted]	[redacted]			
Probabilistic ERG base-case for R^2 versus R-mono					
R^2	[redacted]	[redacted]	[redacted]	[redacted]	£29,010
R-mono	[redacted]	[redacted]			
ERG = Evidence Review Group = ICER = incremental cost effectiveness ratio; QALY = quality-adjusted life year					

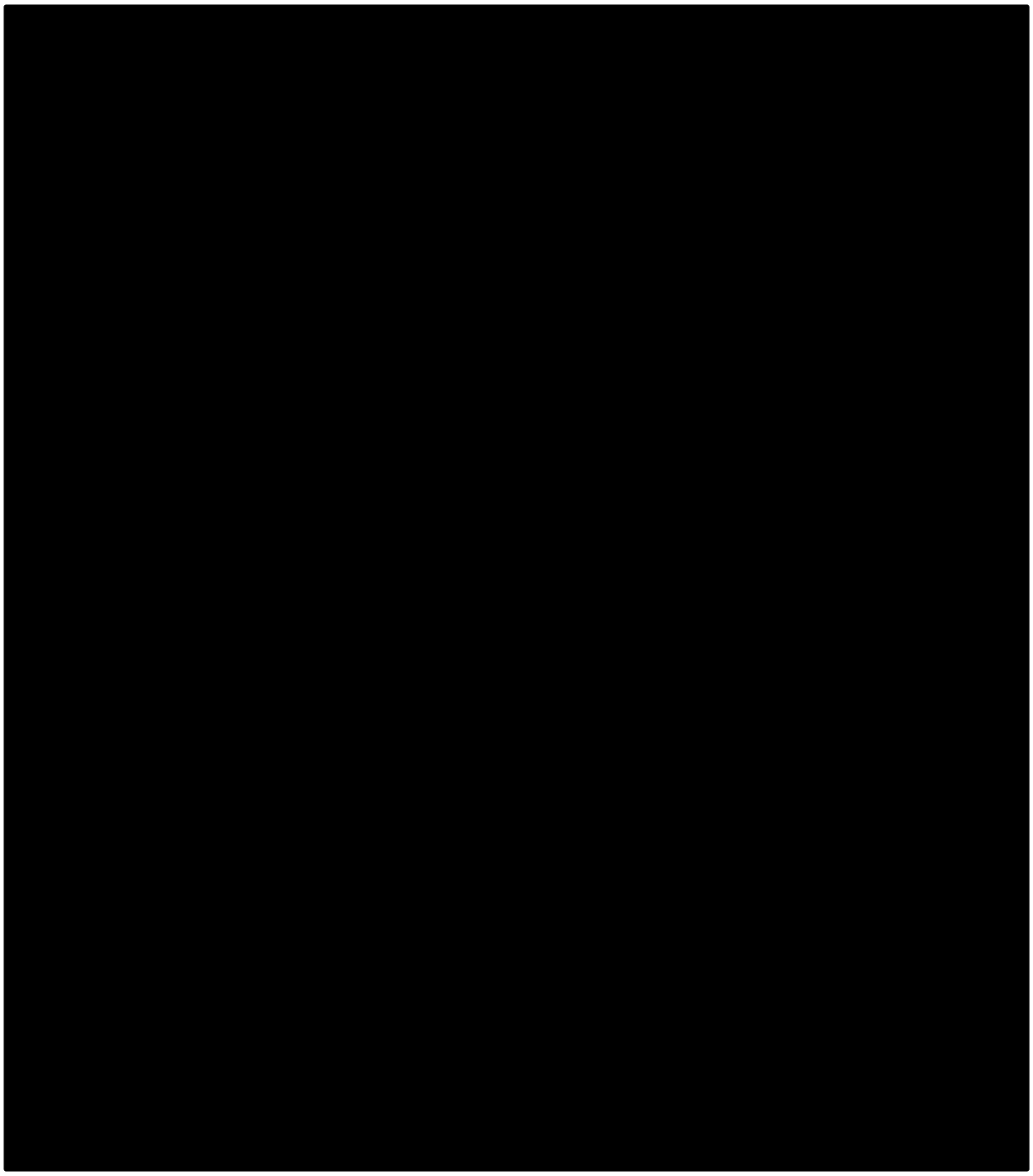


Figure 5.4: Cost effectiveness planes (1,000 iterations) for company and ERG base-case

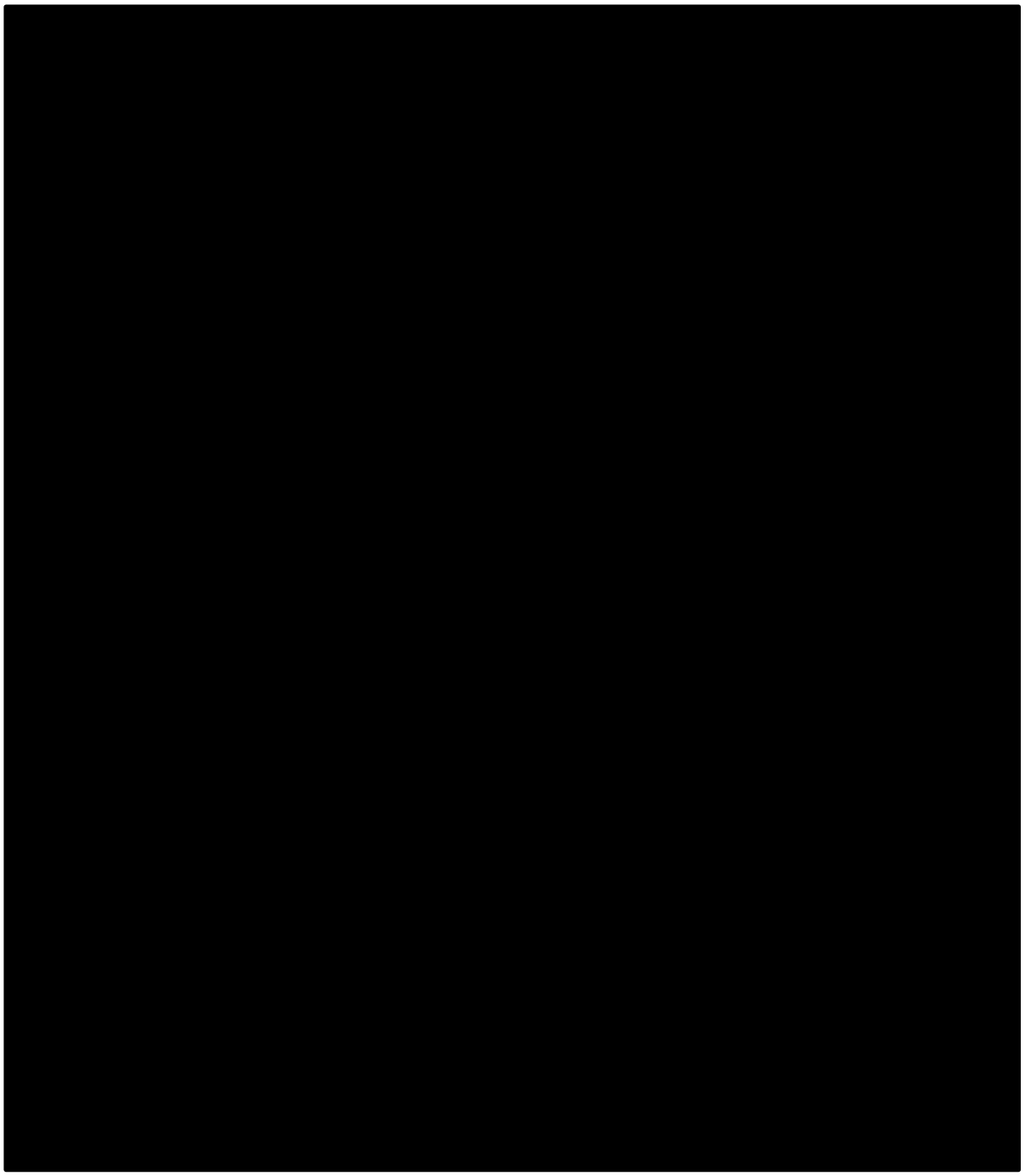


Figure 5.5: ERG base-case cost effectiveness acceptability curve for R^2 versus R-CHOP

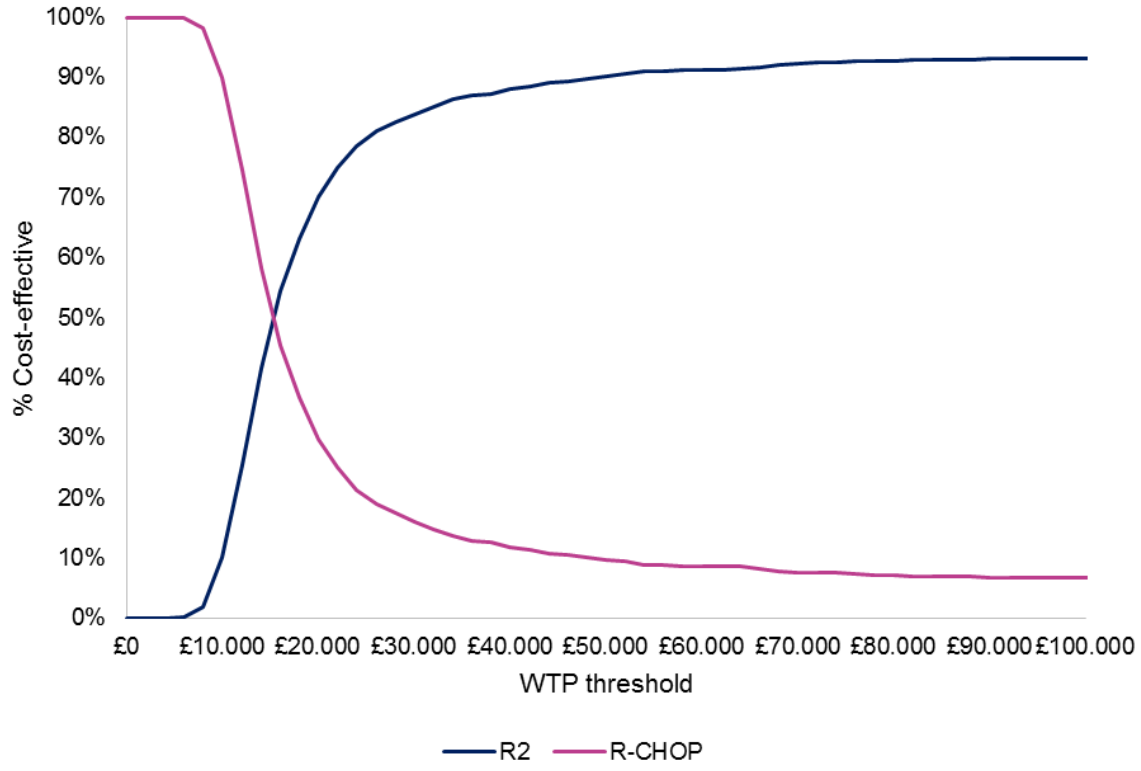


Figure 5.6: ERG base-case cost effectiveness acceptability curve for R^2 versus R-CVP

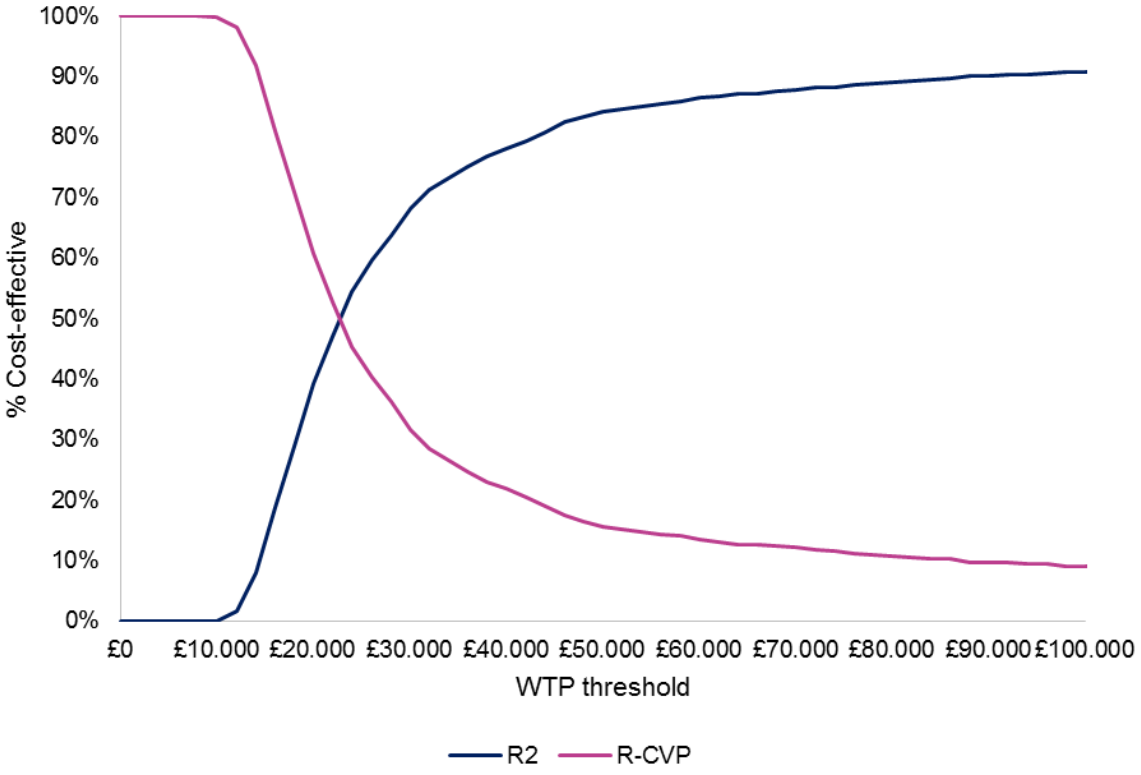
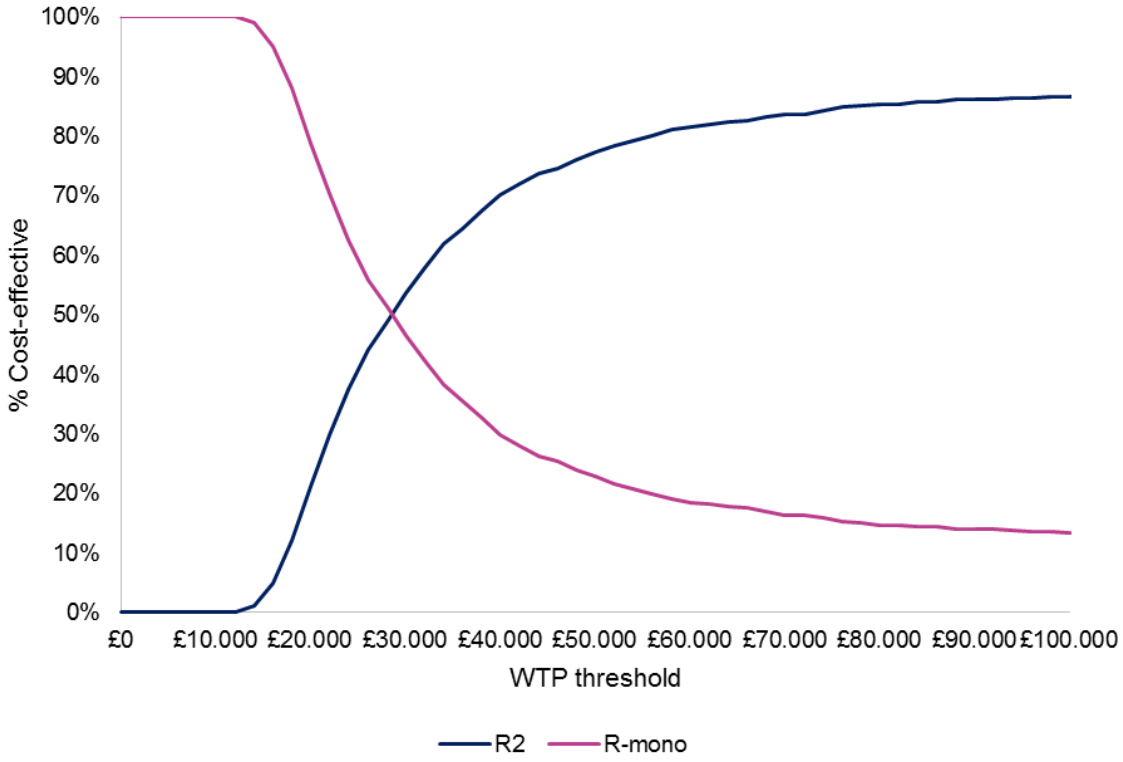


Figure 5.7: ERG base-case cost effectiveness acceptability curve for R^2 versus R-mono



5.3.2 Additional exploratory analyses performed based on the ERG base-case

Additional sensitivity analyses were performed to examine the potential impact of alternative assumptions on the cost effectiveness estimates. These were all performed using the ERG base-case. Results are presented in Table 6.2 in Section 6 of this report.

Exploratory analyses using the ERG base-case:

1. Alternative PFS distributions: use Weibull for PFS both arms (for the R-mono comparison, generalised gamma was used as the alternative PFS distribution) (section 5.2.6)
2. Alternative PFS distributions: use exponential For PFS R^2 and Weibull for PFS comparator (not applied to R-mono comparison) (section 5.2.6)
3. Treatment waning effect after three-year cut-off (section 5.2.6)
4. Treatment waning effect after seven-year cut-off (section 5.2.6)
5. Adverse events for comparator taken from van Oers et al. (2006)[39] (Not applicable in R-mono comparison) (section 5.2.7)
6. FL-only population (section 5.2.3)
7. Apply same subsequent treatment costs for R^2 as for R-CHOP/R-CVP (Not applicable in R-mono comparison) (section 5.2.9)
8. Alternative utilities taken from Wild et al. (2006)[70] 0.805 for PF, 0.736 for PP off treatment, and 0.62 for PP on treatment (section 5.2.8)
9. Source for R-CHOP efficacy taken from van Oers et al. (Not applicable in R-mono comparison) (section 5.2.6)
10. Alternative utilities taken for PP states taken from Pereira et al. (2010)[95] 0.45 for both PP states. (section 5.2.8)

5.3.3 Subgroup analyses performed based on the ERG base-case

No subgroup analyses were performed.

5.4 Conclusions of the cost effectiveness section

Separate sets of searches were conducted to identify cost effectiveness studies, health-related quality of life studies and healthcare resource use evidence. The CS provided clear, transparent and reproducible searches. A good range of databases and additional resources were searched.

The company submission was largely in line with the NICE reference case. The CS partly deviated from the scope, however, where it concerned the comparators modelled. More specifically, R-mono was excluded while direct evidence existed for R^2 versus R-mono, and in the refractory population O-Benda was the sole comparator while NICE had explicitly stated it was not a relevant comparator for this appraisal.

The ERG had concerns about the appropriateness of the partitioned survival model approach and its superiority over a state transition model and would have liked to see both approaches properly explored.

The ERG was concerned about the company pooling MZL and FL populations in the model, assuming they are similar. The ICER for the company's FL-only scenario was substantially higher for the R-CHOP and R-CVP comparisons. This raises serious doubts about the validity of this assumption, and the ERG considered this to be a relevant source of uncertainty.

The main concern of the ERG was the questionable trustworthiness of R^2 efficacy resulting from the indirect comparison, which seemed to be inflated relative to the direct comparison data from AUGMENT. Although the ERG did not have the necessary data to quantify this uncertainty, it may have lowered the ICER substantially.

The ERG had concerns about the way survival curves were selected. The choice of OS curve was mainly based on a previous STA. In particular the choice of PFS curves was not sufficiently justified and appeared sub-optimal, with a likely overestimation of PFS in the R^2 arm, and substantial underestimation of PFS in the first year for R-CHOP and R-CVP. This matter was exacerbated by the high utility values for all health states. The ERG considered these to be potentially overestimated, being higher than or comparable to those in the general population. With utilities remaining high throughout the model, any adjustment in survival curves only had little impact on the ICER, as a high utility post-progression implied there was hardly any penalty on progression in terms of cost effectiveness.

The ERG considered the source used to inform the model concerning AE incidences for R-CHOP and R-CVP to be likely biased, being based on a previously untreated population.

With respect to costs and resource use, the ERG considered the costs of subsequent treatment for R-CHOP and R-CVP to be likely overestimated, as they were based on a mixed R-chemo population from HMRN, while also data specific to R-CHOP and R-CVP separately were available from this source. This was adjusted for in the ERG base-case. The ERG was also concerned about the fact that in the post-progression on treatment phase, there would be a one-off cost for subsequent treatments only, which may be not be reflective of the long-term situation in this health state. As patients in the R^2 arm remain in this health state for a longer time on average, applying costs as one-off possibly favoured R^2.

The ERG made various adjustments to the company base-case. The probabilistic ERG base-case for R^2 versus R-CHOP was £15,818 per QALY gained (based on 1,000 iterations). For R^2 versus R-CVP, the ICER was £23,367 and for R^2 versus R-mono, it was £29,010.

Deterministic scenario analyses were performed to examine the potential impact of alternative assumptions on the cost-effectiveness estimates. For the R-CHOP/R-CVP comparisons, using R-CHOP and R-CVP efficacy from van Oers et al. would lower the ICER substantially, £8,251 for R^2 versus R-

CHOP and £13,315 for R^2 versus R-CVP. Alternative assumptions regarding lowered utilities in the PP health states had the most significant upward impact, increasing the ICER to £33,626 for R^2 versus R-CHOP and £47,281 for R^2 versus R-CVP. For the R-mono comparison, lowering the PP health state utility had the opposite effect, lowering the ICER to £17,826. Another influential scenario was the change of time-point where treatment waning starts to three years (instead of five years in base-case). This increased the ICER to £40,543.

In conclusion, even though the ERG base-case ICER for R-CHOP was below £20,000, the uncertainty around the cost effectiveness of R^2 is substantial, mainly caused by the possible bias introduced by the indirect treatment comparison, which could not be accounted for in the ERG analyses. The ICER for R-CVP is higher and suffers from the same uncertainty. The R-mono analysis is based on a direct comparison, but is also surrounded by substantial uncertainty, as the ICER is rather sensitive to, for instance, the time-point at which treatment waning starts and utilities in the PP health state.

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

6.1 Analyses undertaken by the ERG

In Section 5.3 the ERG base-case was presented, which was based on various changes compared to the company base-case. Tables 6.1 to 6.3 show how individual changes impact the results plus the combined effect of all changes simultaneously, for the R-CHOP, R-CVP, and R-mono comparators, respectively. The exploratory scenario analyses are presented in Tables 6.4 to 6.6 respectively. These are all conditional on the ERG base-case. The analyses numbers in Tables 6.1 to 6.6 correspond to the analyses numbers reported in Section 5.3 of this report. The submitted model files contain technical details on the analyses performed by the ERG (e.g. the "ERG" sheet provides an overview of the cells that were altered for each adjustment). The ERG wishes to emphasise that all ERG analyses (except the R-mono comparison) are conditional upon the MAIC results for which uncertainty could not be quantified or incorporated in the economic model.

Although the tables below report pairwise comparisons only, R-CHOP and R-CVP could also be compared to R^2 in a fully incremental analysis. However, as R-CHOP and R-CVP are by assumption equally effective, and R-CHOP is always the more costly strategy given the higher rate of ASCT performed in the R-CHOP patient population, it is not to be expected that there will be any shifts in the relative comparisons within the fully incremental analysis. Therefore, in practice, the relevant comparison will be R^2 versus R-CVP. For R-mono, a full incremental analysis on the scenarios is not applicable, as a different set of scenarios was performed here, and, more importantly, because costs and QALYs in R^2 are different in this comparison.

Table 6.1: Deterministic ERG base-case for R2 versus R-CHOP comparison

Technologies	Total costs	Total QALYs	Incremental costs	Incremental QALYs	ICER (£/QALY)
CS original base-case					
R^2	[redacted]	[redacted]	[redacted]	[redacted]	£11,471
R-CHOP	[redacted]	[redacted]			
Fixing violations (1, include AEs related to subs Tx in R^2)					
R^2	[redacted]	[redacted]	[redacted]	[redacted]	£11,544
R-CHOP	[redacted]	[redacted]			
Fixing violations (2, use pooled R-CHOP/R-CVP subs Tx instead of mixed R-chemo)					
R^2	[redacted]	[redacted]	[redacted]	[redacted]	£12,206
R-CHOP	[redacted]	[redacted]			
Fixing violations (3, cap utilities at the general population level)					
R^2	[redacted]	[redacted]	[redacted]	[redacted]	£11,977
R-CHOP	[redacted]	[redacted]			
Matter of judgement (4, use exponential for OS in both arms)					
R^2	[redacted]	[redacted]	[redacted]	[redacted]	£12,345
R-CHOP	[redacted]	[redacted]			
Matter of judgement (5, use log-logistic for PFS in R^2 and Weibull for PFS comparator)					
R^2	[redacted]	[redacted]	[redacted]	[redacted]	£13,429
R-CHOP	[redacted]	[redacted]			

Matter of judgement (6, use log-logistic for TTNLT both arms)					
R^2	[redacted]	[redacted]	[redacted]	[redacted]	£11,484
R-CHOP	[redacted]	[redacted]			
ERG base-case (deterministic)					
R^2	[redacted]	[redacted]	[redacted]	[redacted]	£15,505
R-CHOP	[redacted]	[redacted]			
ERG base-case (probabilistic)					
R^2	[redacted]	[redacted]	[redacted]	[redacted]	£15,818
R-CHOP	[redacted]	[redacted]			

Table 6.2: Deterministic ERG base-case for R2 versus R-CVP comparison

Technologies	Total costs	Total QALYs	Incremental costs	Incremental QALYs	ICER (£/QALY)
CS original base-case					
R^2	[redacted]	[redacted]	[redacted]	[redacted]	£16,814
R-CVP	[redacted]	[redacted]			
Fixing violations (1, include AEs related to subs Tx in R^2)					
R^2	[redacted]	[redacted]	[redacted]	[redacted]	£16,888
R-CVP	[redacted]	[redacted]			
Fixing violations (2, use pooled R-CHOP/R-CVP subs Tx instead of mixed R-chemo)					
R^2	[redacted]	[redacted]	[redacted]	[redacted]	£17,549
R-CVP	[redacted]	[redacted]			
Fixing violations (3, cap utilities at the general population level)					
R^2	[redacted]	[redacted]	[redacted]	[redacted]	£17,557
R-CVP	[redacted]	[redacted]			
Matter of judgement (4, use exponential for OS in both arms)					
R^2	[redacted]	[redacted]	[redacted]	[redacted]	£18,304
R-CVP	[redacted]	[redacted]			
Matter of judgement (5, use log-logistic for PFS in R^2 and Weibull for PFS compator)					
R^2	[redacted]	[redacted]	[redacted]	[redacted]	£18,875
R-CVP	[redacted]	[redacted]			
Matter of judgement (6, use log-logistic for TTNLT both arms)					
R^2	[redacted]	[redacted]	[redacted]	[redacted]	£16,867
R-CVP	[redacted]	[redacted]			
ERG base-case (deterministic)					
R^2	[redacted]	[redacted]	[redacted]	[redacted]	£21,759
R-CVP	[redacted]	[redacted]			
ERG base-case (probabilistic)					
R^2	[redacted]	[redacted]	[redacted]	[redacted]	£23,367
R-CVP	[redacted]	[redacted]			

Table 6.3: Deterministic ERG base-case for R2 versus R-mono comparison

Technologies	Total costs	Total QALYs	Incremental costs	Incremental QALYs	ICER (£/QALY)
CS original base-case					
R^2	[redacted]	[redacted]	[redacted]	[redacted]	£22,580
R-mono	[redacted]	[redacted]			
Fixing violations (1, include AEs related to subs Tx in R^2)					
R^2	[redacted]	[redacted]	[redacted]	[redacted]	£22,673
R-mono	[redacted]	[redacted]			
Fixing violations (3, cap utilities at the general population level)					
R^2	[redacted]	[redacted]	[redacted]	[redacted]	£24,054
R-mono	[redacted]	[redacted]			
Matter of judgement (4, use exponential for OS both arms)					
R^2	[redacted]	[redacted]	[redacted]	[redacted]	£25,318
R-mono	[redacted]	[redacted]			
Base-case (deterministic)					
R^2	[redacted]	[redacted]	[redacted]	[redacted]	£27,372
R-mono	[redacted]	[redacted]			
Base-case (probabilistic)					
R^2	[redacted]	[redacted]	[redacted]	[redacted]	£29,010
R-mono	[redacted]	[redacted]			

Table 6.4: Deterministic scenario analyses (conditional on ERG base-case) for R2 versus R-CHOP

Technologies	Total costs	Total QALYs	Incremental costs	Incremental QALYs	ICER (£/QALY)
ERG base-case					
R^2	[redacted]	[redacted]	[redacted]	[redacted]	£15,505
R-CHOP	[redacted]	[redacted]			
Use Weibull for PFS both arms					
R^2	[redacted]	[redacted]	[redacted]	[redacted]	£16,632
R-CHOP	[redacted]	[redacted]			
Use exponential For PFS R2 and Weibull for PFS comparator					
R^2	[redacted]	[redacted]	[redacted]	[redacted]	£14,915
R-CHOP	[redacted]	[redacted]			
Treatment waning effect at 3 years					
R^2	[redacted]	[redacted]	[redacted]	[redacted]	£19,018
R-CHOP	[redacted]	[redacted]			
Treatment waning effect at 7 years					
R^2	[redacted]	[redacted]	[redacted]	[redacted]	£13,654
R-CHOP	[redacted]	[redacted]			
Adverse events for comparator taken from publication					
R^2	[redacted]	[redacted]	[redacted]	[redacted]	£18,270

Technologies	Total costs	Total QALYs	Incremental costs	Incremental QALYs	ICER (£/QALY)
R-CHOP	[redacted]	[redacted]			
FL-only population					
R^2	[redacted]	[redacted]	[redacted]	[redacted]	£16,680
R-CHOP	[redacted]	[redacted]			
Apply same subsequent treatment costs					
R^2	[redacted]	[redacted]	[redacted]	[redacted]	£18,640
R-CHOP	[redacted]	[redacted]			
Alternative utilities for PP states from Wild et al. (0.62)					
R^2	[redacted]	[redacted]	[redacted]	[redacted]	£21,526
R-CHOP	[redacted]	[redacted]			
Source for R-CHOP/R-CVP efficacy from van Oers					
R^2	[redacted]	[redacted]	[redacted]	[redacted]	£8,251
R-CHOP	[redacted]	[redacted]			
Alternative utilities for PP states from Pereira et al. (0.45)					
R^2	[redacted]	[redacted]	[redacted]	[redacted]	£33,626
R-CHOP	[redacted]	[redacted]			

Table 6.5: Deterministic scenario analyses (conditional on ERG base-case) for R2 versus R-CVP

Technologies	Total costs	Total QALYs	Incremental costs	Incremental QALYs	ICER (£/QALY)
ERG base-case					
R^2	[redacted]	[redacted]	[redacted]	[redacted]	£21,759
R-CVP	[redacted]	[redacted]			
Use Weibull for PFS both arms					
R^2	[redacted]	[redacted]	[redacted]	[redacted]	£22,887
R-CVP	[redacted]	[redacted]			
Use exponential For PFS R2 and Weibull for PFS comparator					
R^2	[redacted]	[redacted]	[redacted]	[redacted]	£21,167
R-CVP	[redacted]	[redacted]			
Treatment waning effect at 3 years					
R^2	[redacted]	[redacted]	[redacted]	[redacted]	£28,562
R-CVP	[redacted]	[redacted]			
Treatment waning effect at 7 years					
R^2	[redacted]	[redacted]	[redacted]	[redacted]	£18,523
R-CVP	[redacted]	[redacted]			
Adverse events for comparator taken from publication					
R^2	[redacted]	[redacted]	[redacted]	[redacted]	£23,618
R-CVP	[redacted]	[redacted]			
FL-only population					
R^2	[redacted]	[redacted]	[redacted]	[redacted]	£22,841

Technologies	Total costs	Total QALYs	Incremental costs	Incremental QALYs	ICER (£/QALY)
R-CVP	[redacted]	[redacted]			
Apply same subsequent treatment costs					
R^2	[redacted]	[redacted]	[redacted]	[redacted]	£24,899
R-CVP	[redacted]	[redacted]			
Alternative utilities for PP states from Wild et al. (0.62)					
R^2	[redacted]	[redacted]	[redacted]	[redacted]	£30,227
R-CVP	[redacted]	[redacted]			
Source for R-CVP efficacy from van Oers					
R^2	[redacted]	[redacted]	[redacted]	[redacted]	£13,315
R-CVP	[redacted]	[redacted]			
Alternative utilities for PP states from Pereira et al. (0.45)					
R^2	[redacted]	[redacted]	[redacted]	[redacted]	£47,281
R-CVP	[redacted]	[redacted]			

Table 6.6: Deterministic scenario analyses (conditional on ERG base-case) for R2 versus R-mono

Technologies	Total costs	Total QALYs	Incremental costs	Incremental QALYs	ICER (£/QALY)
ERG base-case					
R^2	[redacted]	[redacted]	[redacted]	[redacted]	£27,372
R-mono	[redacted]	[redacted]			
Use Generalised gamma for PFS both arms					
R^2	[redacted]	[redacted]	[redacted]	[redacted]	£28,206
R-mono	[redacted]	[redacted]			
Treatment waning effect at 3 years					
R^2	[redacted]	[redacted]	[redacted]	[redacted]	£40,543
R-mono	[redacted]	[redacted]			
Treatment waning effect at 7 years					
R^2	[redacted]	[redacted]	[redacted]	[redacted]	£22,091
R-mono	[redacted]	[redacted]			
FL-only population					
R^2	[redacted]	[redacted]	[redacted]	[redacted]	£17,936
R-mono	[redacted]	[redacted]			
Apply same subsequent treatment costs					
R^2	[redacted]	[redacted]	[redacted]	[redacted]	£30,263
R-mono	[redacted]	[redacted]			
Alternative utilities for PP states from Wild et al. (0.62)					
R^2	[redacted]	[redacted]	[redacted]	[redacted]	£21,349
R-mono	[redacted]	[redacted]			
Alternative utilities for PP states from Pereira et al. (0.45)					
R^2	[redacted]	[redacted]	[redacted]	[redacted]	£17,826

Technologies	Total costs	Total QALYs	Incremental costs	Incremental QALYs	ICER (£/QALY)
R-mono	[redacted]	[redacted]			

7. REFERENCES

[1] Celgene Ltd. *Lenalidomide with rituximab for treated follicular lymphoma and marginal zone lymphoma [ID1374]. Document B. Company evidence submission to National Institute for Health and Care Excellence. Single technology appraisal (STA)*: Celgene Ltd, 2019

[2] Haematological Malignancy Research Network (HMRN). Incidence [Internet]. York: Haematological Malignancy Research Network, Epidemiology & Statistics Group (ECSG) at the University of York, 2019 [accessed 1.5.19]. Available from: https://www.hmrn.org/statistics/incidence

[3] Haematological Malignancy Research Network (HMRN). *Patient's age and treatment for haematological malignancy: a report from the Haematological Malignancy Research Network (HMRN) [Internet]*. York: Haematological Malignancy Research Network, 2014 [accessed 1.5.19] Available from: https://www.hmrn.org/Download.aspx?target=document&id=41

[4] Haematological Malignancy Research Network (HMRN). Survival: follicular lymphoma [Internet]. York: Haematological Malignancy Research Network, Epidemiology & Statistics Group (ECSG) at the University of York, 2019 [accessed 1.5.19]. Available from: https://www.hmrn.org/statistics/survival

[5] Cancer Research UK. Non-Hodgkin lymphoma: survival [Internet]. Cancer Research UK, 2018 [accessed 1.5.19]. Available from: https://about-cancer.cancerresearchuk.org/about-cancer/non-hodgkin-lymphoma/survival

[6] Provencio M, Sabin P, Gomez-Codina J, Torrente M, Calvo V, Llanos M, et al. Impact of treatment in long-term survival patients with follicular lymphoma: a Spanish Lymphoma Oncology Group registry. *PLoS One* 2017;12(5):e0177204.

[7] Johnson PW, Rohatiner AZ, Whelan JS, Price CG, Love S, Lim J, et al. Patterns of survival in patients with recurrent follicular lymphoma: a 20-year study from a single center. *J Clin Oncol* 1995;13(1):140-7.

[8] Pettengell R, Donatti C, Hoskin P, Poynton C, Kettle PJ, Hancock B, et al. The impact of follicular lymphoma on health-related quality of life. *Ann Oncol* 2008;19(3):570-6.

[9] El-Zimaity M. Marginal zone B-cell lymphoma [Internet]. Medscape, 2014 [accessed 1.5.19]. Available from: https://emedicine.medscape.com/article/1610599-overview

[10] Olszewski AJ, Castillo JJ. Survival of patients with marginal zone lymphoma: analysis of the Surveillance, Epidemiology, and End Results database. *Cancer* 2013;119(3):629-38.

[11] Denlinger NM, Epperla N, William BM. Management of relapsed/refractory marginal zone lymphoma: focus on ibrutinib. *Cancer Manag Res* 2018;10:615-24.

[12] Office of National Statistics (ONS). Cancer Registration Statistics, England, 2017. (Updated: 26 April 2019) [Internet]. Office of National Statistics (ONS), 2019 [accessed 1.5.19]. Available from: https://www.ons.gov.uk/file?uri=/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/datasets/cancerregistrationstatisticscancerregistrationstatisticsengland/2017/finalreferencetables.xlsx

[13] Wang HI, Roman E, Crouch S, Aas E, Burton C, Patmore R, et al. A generic model for follicular lymphoma: predicting cost, life expectancy, and quality-adjusted-life-year using UK population-based observational data. *Value Health* 2018;21(10):1176-85.

[14] Lymphoma Action. MALT lymphoma (extranodal marginal zone lymphoma) [Internet]. Lymphoma Action, 2018 [accessed 1.5.19]. Available from: https://lymphoma-action.org.uk/sites/default/files/media/documents/2018-08/LYMweb0218MALT2017v1.2.pdf

[15] Lymphoma Action. Splenic marginal zone lymphoma [Internet]. Lymphoma Action, 2018 [accessed 1.5.19]. Available from: https://lymphoma-action.org.uk/sites/default/files/media/documents/2018-06/LYMweb0220SplenicMZ2017v1.2.pdf

[16] Lymphoma Action. Nodal marginal zone lymphoma [Internet]. Lymphoma Action, 2019 [accessed 1.5.19]. Available from: https://lymphoma-action.org.uk/sites/default/files/media/documents/2018-06/LYMweb0219NodalMZ2017v1.2.pdf

[17] Chiu H, Trisal P, Bjorklund C, Carrancio S, Torano EG, Guarinos C, et al. Combination lenalidomide-rituximab immunotherapy activates anti-tumour immunity and induces tumour cell death by complementary mechanisms of action in follicular lymphoma. *Br J Haematol* 2019;185(2):240-253.

[18] Celgene. *Draft summary of product characteristics. Revlimid® (Lenalidomide)*: Celgene, 2019

[19] Celgene. Celgene Indolent Non-Hodgkin Lymphoma Advisory Board Meeting. (Final meeting report: NP-UK-REV-0086) March 2019. (Updated: August 2019) Data on file. 2019.

[20] National Institute for Health and Care Excellence. *Obinutuzumab with bendamustine for treating follicular lymphoma refractory to rituximab. NICE technology appraisal guidance 472 [Internet]*. London: National Institute for Health and Care Excellence, 2017 [accessed 5.1.19] Available from: https://www.nice.org.uk/guidance/ta472

[21] National Institute for Health and Care Excellence. *Rituximab for the first-line treatment of stage III-IV follicular lymphoma. NICE technology appraisal guidance 243 [Internet]*. London: National Institute for Health and Care Excellence, 2012 [accessed 26.4.19] Available from: https://www.nice.org.uk/guidance/ta243

[22] National Institute for Health and Care Excellence. *Rituximab for the first-line maintenance treatment of follicular non-Hodgkin's lymphoma. NICE technology appraisal guidance 226 [Internet]*. London: National Institute for Health and Care Excellence, 2011 [accessed 5.5.19] Available from: https://www.nice.org.uk/guidance/ta226

[23] National Institute for Health and Care Excellence. *Obinutuzumab for untreated advanced follicular lymphoma. NICE technology appraisal guidance 513 [Internet]*. London: National Institute for Health and Care Excellence, 2018 [accessed 5.1.19] Available from: https://www.nice.org.uk/guidance/ta513

[24] National Institute for Health and Care Excellence. NICE Pathways: Treating follicular lymphoma [Internet]. London: National Institute for Health and Care Excellence, Updated Jan 2019 [accessed 13.2.19]. Available from: https://pathways.nice.org.uk/pathways/non-hodgkins-lymphoma/treating-follicular-lymphoma#content=view-node%3Anodes-first-line-treatment-for-early-stage-disease

[25] National Institute for Health and Care Excellence. *Non-Hodgkin's lymphoma: diagnosis and management. NICE guideline 52 [Internet]*. London: National Institute for Health and Care Excellence, 2016 [accessed 25.4.19] Available from: https://www.nice.org.uk/guidance/ng52

[26] Celgene Ltd. *Lenalidomide for treated follicular lymphoma and marginal zone lymphoma [ID1374]. Response to request for clarification from the ERG. August 2019*: Celgene Ltd, 2019

[27] National Institute for Health and Care Excellence. Cancer Drugs Fund [Internet]. London: National Institute for Health and Care Excellence, [accessed 3.9.19]. Available from: https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/nice-technology-appraisal-guidance/cancer-drugs-fund

[28] National Institute for Health and Care Excellence. *Lenalidomide with rituximab for previously treated follicular lymphoma and marginal zone lymphoma. Response to consultee and commentator comments on the draft remit and draft scope (pre-referral) [Internet]*. London: National Institute for Health and Care Excellence, 2019 [accessed 25.7.19] Available from:

https://www.nice.org.uk/guidance/gid-ta10323/documents/scope-consultation-comments-and-responses

[29] Dreyling M, Thieblemont C, Gallamini A, Arcaini L, Campo E, Hermine O, et al. ESMO Consensus conferences: guidelines on malignant lymphoma. Part 2: marginal zone lymphoma, mantle cell lymphoma, peripheral T-cell lymphoma. *Ann Oncol* 2013;24(4):857-77.

[30] Lymphoma Research Foundation. *Getting the facts. Marginal zone lymphoma [Internet]*: Lymphoma Research Foundation, 2018 [accessed 5.5.19] Available from: https://www.lymphoma.org/wp-content/uploads/2018/07/MarginalZoneLymphomaF_Fact-Sheet_2018.pdf

[31] Celgene Ltd. [redacted] clinical validation. 12 June 2019. Data on file. 2019.

[32] National Institute for Health and Care Excellence. *Lenalidomide with rituximab for previously treated follicular lymphoma and marginal zone lymphoma. Final scope [Internet]*. London: National Institute for Health and Care Excellence, 2019 [accessed 25.7.19] Available from: https://www.nice.org.uk/guidance/gid-ta10323/documents/final-scope

[33] Celgene Ltd. *Lenalidomide with rituximab for treated follicular lymphoma and marginal zone lymphoma [ID1374]. Document A. Company evidence submission summary for committee. Single technology appraisal (STA)*: Celgene Ltd, 2019

[34] Celgene Ltd. *Lenalidomide with rituximab for treated follicular lymphoma and marginal zone lymphoma [ID1374]. Document B. Appendices. Single technology appraisal (STA)*: Celgene Ltd, 2019

[35] Leonard JP, Trneny M, Izutsu K, Fowler NH, Hong X, Zhu J, et al. AUGMENT: a phase III study of lenalidomide plus rituximab versus placebo plus rituximab in relapsed or refractory indolent lymphoma. *J Clin Oncol* 2019;37(14):1188-99.

[36] Andorsky D, Coleman M, Yacoub A, Melear JM, Brooks HD, Fanning SR, et al. Response rate to lenalidomide plus rituximab (R2) as independent of number of prior lines of therapy: interim analysis of initial phase of MAGNIFY phase IIIb study of R2 followed by maintenance in relapsed/refractory indolent NHL. *J Clin Oncol* 2018;36(15 Suppl):7516.

[37] Leonard JP, Jung SH, Johnson J, Pitcher BN, Bartlett NL, Blum KA, et al. Randomized trial of lenalidomide alone versus lenalidomide plus rituximab in patients with recurrent follicular lymphoma: CALGB 50401 (ALLIANCE). *J Clin Oncol* 2015;33(31):3635-40.

[38] Tuscano JM, Dutia M, Chee K, Brunson A, Reed-Pease C, Abedi M, et al. Lenalidomide plus rituximab can produce durable clinical responses in patients with relapsed or refractory, indolent non-Hodgkin lymphoma. *Br J Haematol* 2014;165(3):375-81.

[39] van Oers MH, Klasa R, Marcus RE, Wolf M, Kimby E, Gascoyne RD, et al. Rituximab maintenance improves clinical outcome of relapsed/resistant follicular non-Hodgkin lymphoma in patients both with and without rituximab during induction: results of a prospective randomized phase 3 intergroup trial. *Blood* 2006;108(10):3295-301.

[40] Sehn LH, Chua N, Mayer J, Dueck G, Trněný M, Bouabdallah K, et al. Obinutuzumab plus bendamustine versus bendamustine monotherapy in patients with rituximab-refractory indolent non-Hodgkin lymphoma (GADOLIN): a randomised, controlled, open-label, multicentre, phase 3 trial. *Lancet Oncol* 2016;17(8):1081-93.

[41] Phillippo D, Ades AE, Dias S, Palmer S, Abrams KR, Welton NJ. *NICE DSU technical support document 18: methods for population-adjusted indirect comparisons in submissions to NICE [Internet]*. Sheffield: NICE Decision Support Unit, 2016 [accessed 2.5.19] Available from: http://www.nicedsu.org.uk

[42] Guyot P, Ades AE, Ouwens MJ, Welton NJ. Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan-Meier survival curves. *BMC Med Res Methodol* 2012;12:9.

[43] Guzauskas GF, Masaquel A, Reyes C, Bernaards C, Krivasi T, Veenstra DL. Cost-effectiveness of obinutuzumab plus bendamustine followed by obinutuzumab monotherapy for the treatment of follicular lymphoma patients who relapse after or are refractory to a rituximab-containing regimen in the US. *J Med Econ* 2018;21(10):960-7.

[44] Haukaas FS, Ohna A, Krivasi T. Cost-effectiveness of obinutuzumab in combination with bendamustine followed by obinutuzumab maintenance versus bendamustine alone in treatment of patients with rituximab-refractory follicular lymphoma in Norway. *Appl Health Econ Health Policy* 2018;16(4):569-77.

[45] All Wales Therapeutics & Toxicology Centre. *AWMSG Secretariat Assessment Report. Idelalisib (Zydelig®) 100 mg and 150 mg film-coated tablets. Reference number: 2597 [Internet]*: All Wales Therapeutics & Toxicology Centre, 2017 [accessed 14.3.19] Available from: http://www.awmsg.org/awmsgonline/app/appraisalinfo/2597

[46] Scottish Medicines Consortium (SMC). *Idelalisib, 100mg and 150mg film-coated tablets (Zydelig®). SMC No. (1039/15) [Internet]*. Glasgow: Scottish Medicines Consortium (SMC), 2015 [accessed 14.3.19] Available from: https://www.scottishmedicines.org.uk/

[47] Blommestein HM, Issa DE, Pompen M, Ten Hoor G, Hogendoorn M, Joosten P, et al. Cost-effectiveness of rituximab as maintenance treatment for relapsed follicular lymphoma: results of a population-based study. *Eur J Haematol* 2014;92(5):398-406.

[48] Pink J, Lane S, Hughes DA. Mechanism-based approach to the economic evaluation of pharmaceuticals: pharmacokinetic/pharmacodynamic/pharmacoeconomic analysis of rituximab for follicular lymphoma. *Pharmacoeconomics* 2012;30(5):413-29.

[49] Vandekerckhove S, Lamotte M, Iannazzo S, Chirita O. Cost utility analysis of bortezomib in a biomarker positive subgroup of relapsed and refractory folicular lymphoma. *Value Health* 2012;15(7):A427.

[50] Deconinck E, Miadi-Fargier H, Pen CL, Brice P. Cost effectiveness of rituximab maintenance therapy in follicular lymphoma: long-term economic evaluation. *Pharmacoeconomics* 2010;28(1):35-46.

[51] Papadakis K, Follows GA, Boyer J, Bashir Z, Ball P, Aultman R, et al. Cost effectiveness analysis of rituximab maintenance in patients with untreated high tumour burden follicular lymphoma after response to immunochemotherapy: a UK national healthcare services perspective. *Blood* 2010;116(21):3833.

[52] Follicular Lymphoma Pharmacoeconomics Group. [Rituximab cost analysis for maintenance treatment of patients with follicular lymphoma]. *Farm Hosp* 2008;32(1):25-34.

[53] Chiattone C, Saggia MG, Santos EV. Cost-utility analysis of maintenance therapy with rituximab for follicular lymphoma in Brazil according to local HTA guidelines. *Ann Oncol* 2008;19(Suppl 8):737P.

[54] Hayslip JW, Simpson KN. Cost-effectiveness of extended adjuvant rituximab for US patients aged 65-70 years with follicular lymphoma in second remission. *Clin Lymphoma Myeloma* 2008;8(3):166-70.

[55] Hornberger J, Reyes C, Lubeck D, Valente N. Economic evaluation of rituximab plus cyclophosphamide, vincristine and prednisolone for advanced follicular lymphoma. *Leuk Lymphoma* 2008;49(2):227-36.

[56] Kasteng F, Erlanson M, Hagberg H, Kimby E, Relander T, Lundkvist J. Cost-effectiveness of maintenance rituximab treatment after second line therapy in patients with follicular lymphoma in Sweden. *Acta Oncol* 2008;47(6):1029-36.

[57] Scottish Medicines Consortium (SMC). *Rituximab 10mg/ml concentrate for infusion (MabThera®) (No.330/06) [Internet]*. Glasgow: Scottish Medicines Consortium (SMC), 2006 [accessed 14.3.19] Available from: https://www.scottishmedicines.org.uk/

[58] Scottish Medicines Consortium (SMC). *Ibritumomab tiuxetan (Zevalin®) No. (171/05). Re-submission [Internet]*. Glasgow: Scottish Medicines Consortium (SMC), 2007 [accessed 14.3.19] Available from: https://www.scottishmedicines.org.uk/

[59] National Institute for Health and Care Excellence. *Rituximab for the treatment of relapsed or refractory stage III or IV follicular non-Hodgkin's lymphoma. NICE technology appraisal guidance 137 [Internet]*. London: National Institute for Health and Care Excellence, 2008 [accessed 1.2.19] Available from: https://www.nice.org.uk/guidance/ta137

[60] Guzauskas GF, Masaquel A, Thuresson P-O, Li J, Reyes CM, Veenstra D. Is obinutuzumab plus chemotherapy followed by obinutuzumab monotherapy for previously untreated follicular lymphoma patients cost-effective in the US? *Blood* 2017;130:5603.

[61] Scottish Medicines Consortium (SMC). Obinutuzumab 1,000mg concentrate for solution for infusion (Gazyvaro®) [Internet]. Glasgow: Scottish Medicines Consortium (SMC), 2017 [accessed 14.3.19]. Available from: https://www.scottishmedicines.org.uk/

[62] Soini EJ, Martikainen JA, Nousiainen T. Treatment of follicular non-Hodgkin's lymphoma with or without rituximab: cost-effectiveness and value of information based on a 5-year follow-up. *Ann Oncol* 2011;22(5):1189-97.

[63] Woods B, Sideris E, Palmer S, Latimer N, Soares M. *NICE DSU technical support document 19: partitioned survival analysis for decision modelling in health care: a critical review [Internet]*. Sheffield: NICE Decision Support Unit, 2017 [accessed 28.5.19] Available from: http://www.nicedsu.org.uk

[64] Leonard JP, Trneny M, Izutsu K, et al. AUGMENT phase III study: Lenalidomide/rituximab (R2) improved efficacy over rituximab/placebo in relapsed/refractory follicular patients irrespective of POD24 status. Paper submitted to 15-ICML International Conference on Malignant Lymphoma; 18-22 Jun 2019; Lugano: Switzerland. 2019.

[65] Haematological Malignancy Research Network (HMRN). Clinical management and outcome of follicular lymphoma (FL) with a focus on relapsed/refractory disease (draft report). 17 July 2019. Haematological Malignancy Research Network, 2019.

[66] Morschhauser F, Fowler NH, Feugier P, Bouabdallah R, Tilly H, Palomba ML, et al. Rituximab plus lenalidomide in advanced untreated follicular lymphoma. *N Engl J Med* 2018;379(10):934-47.

[67] van Oers MH, Van Glabbeke M, Giurgea L, Klasa R, Marcus RE, Wolf M, et al. Rituximab maintenance treatment of relapsed/resistant follicular non-Hodgkin's lymphoma: long-term outcome of the EORTC 20981 phase III randomized intergroup study. *J Clin Oncol* 2010;28(17):2853-8.

[68] National Institute for Health and Care Excellence. *Non-Hodgkin's lymphoma: diagnosis and management. NICE guideline 52 [Internet]*. London: National Institute for Health and Care Excellence, 2016 [accessed 25.7.19] Available from: https://www.nice.org.uk/guidance/ng52

[69] Leger C, Sabloff M, McDiarmid S, Bence-Bruckler I, Atkins H, Bredeson C, et al. Outpatient autologous hematopoietic stem cell transplantation for patients with relapsed follicular lymphoma. *Ann Hematol* 2006;85(10):723-9.

[70] Wild D, Walker M, Pettengell R, Lewis G. Pcn62 utility elicitation in patients with follicular lymphoma. *Value Health* 2006;9(6):A294.

[71] Pereira C, Negreiro F, Silva C. PSY43 Economic analysis of rituximab in combination with cyclophosphamide, vincristine and prednisolone in the treatment of patients with advanced follicular lymphoma in Portugal. *Value Health* 2010;13(7):A468.

[72] Nafees B, Stafford M, Gavriel S, Bhalla S, Watkins J. Health state utilities for non small cell lung cancer. *Health Qual Life Outcomes* 2008;6:84.

[73] National Institute for Health and Care Excellence. *Pixantrone monotherapy for treating multiply relapsed or refractory aggressive non-Hodgkin's B-cell lymphoma. NICE technology appraisal guidance 306 [Internet]*. London: National Institute for Health and Care Excellence, 2014 [accessed 25.4.19] Available from: https://www.nice.org.uk/guidance/ta306

[74] Swinburn P, Lloyd A, Nathan P, Choueiri TK, Cella D, Neary MP. Elicitation of health state utilities in metastatic renal cell carcinoma. *Curr Med Res Opin* 2010;26(5):1091-6.

[75] Beusterien KM, Davies J, Leach M, Meiklejohn D, Grinspan JL, O'Toole A, et al. Population preference values for treatment outcomes in chronic lymphocytic leukaemia: a cross-sectional utility study. *Health Qual Life Outcomes* 2010;8:50.

[76] Stein EM, Yang M, Guerin A, Gao W, Galebach P, Xiang CQ, et al. Assessing utility values for treatment-related health states of acute myeloid leukemia in the United States. *Health Qual Life Outcomes* 2018;16(1):193.

[77] Tolley K, Goad C, Yi Y, Maroudas P, Haiderali A, Thompson G. Utility elicitation study in the UK general public for late-stage chronic lymphocytic leukaemia. *Eur J Health Econ* 2013;14(5):749-59.

[78] National Institute for Health and Care Excellence. *Eribulin for treating locally advanced or metastatic breast cancer after 2 or more chemotherapy regimens. NICE technology appraisal guidance 423 [Internet]*. London: National Institute for Health and Care Excellence, 2016 [accessed 25.4.19] Available from: https://www.nice.org.uk/guidance/ta423

[79] Hannouf MB, Sehgal C, Cao JQ, Mocanu JD, Winquist E, Zaric GS. Cost-effectiveness of adding cetuximab to platinum-based chemotherapy for first-line treatment of recurrent or metastatic head and neck cancer. *PLoS One* 2012;7(6):e38557.

[80] Doyle S, Lloyd A, Walker M. Health state utility scores in advanced non-small cell lung cancer. *Lung Cancer* 2008;62(3):374-80.

[81] Kind P, Hardman G, Macran S. *UK population norms for EQ-5D. Discussion paper 172*. York: Centre for Health Economics, 1999 Available from: https://www.york.ac.uk/media/che/documents/papers/discussionpapers/CHE%20Discussion%20Paper%20172.pdf

[82] NHS Improvement. *NHS reference costs 2017/18 [Internet]*. London: NHS Improvement, 2018 [accessed 26.4.19] Available from: https://improvement.nhs.uk/resources/reference-costs/

[83] Personal Social Services Research Unit. *Unit costs of health and social care 2018*. Canterbury: University of Kent, 2018 [accessed 25.7.19] Available from: https://www.pssru.ac.uk/project-pages/unit-costs/unit-costs-2018/

[84] MIMS. MabThera [Internet]. 2019 [accessed 27.2.19]. Available from: https://www.mims.co.uk/drugs/cancer/antineoplastics/mabthera

[85] MIMS. Revlimid [Internet]. 2019 [accessed 25.4.19]. Available from: https://www.mims.co.uk/drugs/cancer/antineoplastics/revlimid

[86] MIMS. Neupogen [Internet]. 2019 [accessed 31.5.19]. Available from: https://www.mims.co.uk/drugs/cardiovascular-system/anaemia-neutropenia/neupogen

[87] MIMS. Allopurinol [Internet]. 2019 [accessed 31.5.19]. Available from: https://www.mims.co.uk/drugs/musculoskeletal-disorders/gout/allopurinol

[88] MIMS. Gazyvaro [Internet]. 2019 [accessed 27.2.19]. Available from: https://www.mims.co.uk/drugs/cancer/antineoplastics/gazyvaro

[89] MIMS. Rixathon [Internet]. 2019 [accessed 27.2.19]. Available from: https://www.mims.co.uk/drugs/cancer/antineoplastics/rixathon

[90] MIMS. Zydelig [Internet]. 2019 [accessed 1.6.19]. Available from: https://www.mims.co.uk/drugs/cancer/antineoplastics/zydelig

[91] GOV.UK. Drugs and pharmaceutical electronic market information tool (eMIT) [Internet]. 2019 [accessed 25.4.19]. Available from: https://www.gov.uk/government/publications/drugs-and-pharmaceutical-electronic-market-information-emit

[92] Round J, Jones L, Morris S. Estimating the cost of caring for people with cancer at the end of life: a modelling study. *Palliat Med* 2015;29(10):899-907.

[93] Vemer P, Corro Ramos I, van Voorn GA, Al MJ, Feenstra TL. AdViSHE: a validation-assessment tool of health-economic models for decision makers and model users. *Pharmacoeconomics* 2016;34(4):349-61.

[94] Kaltenthaler E, Carroll C, Hill-McManus D, Scope A, Holmes M, Rice S, et al. The use of exploratory analyses within the National Institute for Health and Care Excellence single technology appraisal process: an evaluation and qualitative analysis. *Health Technol Assess* 2016;20(26):1-48.

[95] Pereira C, Negreiro F, Silva C. Economic analysis of rituximab in combination with cyclophosphamide, vincristine and prednisolone in the treatment of patients with advanced follicular lymphoma in Portugal. PSY43. *Value Health* 2010;13(7):A468.

Appendix 1: Additional results from the AUGMENT trial

Table 4.5 in Section 4.2.5 of this report presents a summary of the main results from the AUGMENT trial for the full ITT population. Results for FL and MZL separately are reported in Tables A1.1 and A1.2 below.

Table A1.1: Summary of results from the AUGMENT trial: ITT population (FL).

Endpoint	FL	
	R[2] (n=147)	**R-mono (n=148)**
Median OS, months (95% CI)[a]	[redacted]	[redacted]
Hazard ratio (95% CI)	[redacted][c]	
Median PFS, months (95% CI)[a]	[redacted]	[redacted]
Hazard ratio (95% CI)	[redacted][c]	
Best response, n (%)		
ORR (CR+PR)	[redacted]	[redacted]
95% CI[d]	[redacted]	[redacted]
p-value	[redacted]	
CR rate	[redacted]	[redacted]
95% CI[d]	[redacted]	[redacted]
p-value	[redacted]	
PR	[redacted]	[redacted]
SD	[redacted]	[redacted]
PD/ death	[redacted]	[redacted]
No evidence of disease	[redacted]	[redacted]
Unknown/ND/Missing	[redacted]	[redacted]
Median TTNLT, months (95% CI)[a]	[redacted]	[redacted]
TTNLT rate at 2 years, % (95% CI)	[redacted]	[redacted]
Hazard ratio (95% CI)	[redacted]	
p-value	[redacted]	
Median EFS, months (95% CI)[a]	[redacted]	[redacted]
Hazard ratio (95% CI)	[redacted]	
p-value	[redacted]	
Median TTNCT, months (95% CI)[a]	NR	NR
TTNCT rate at 2 years, % (95% CI)	NR	NR
Hazard ratio (95% CI)	NR	
p-value	NR	
RTNLT		
ORR, n (% [95% CI][d])	NR	NR
p-value	NR	
CR, n (% [95% CI][d])	NR	NR
p-value	NR	
DCRR, n/N (%)	[redacted]	[redacted]

95% CI[d]	[redacted]
p-value	[redacted]
N, Median DOR, months (95% CI)[a]	[redacted]
Hazard ratio (95% CI)[c]	[redacted]
p-value[e]	[redacted]
N, Median DOCR, months (95% CI)[a]	[redacted]
Hazard ratio (95% CI)[h]	[redacted]
p-value	[redacted]

Source: Response to CL, Table 5, pages 19 and 20.
CI = confidence interval; CR = complete response; DCRR = durable complete response rate, DOCR = duration of complete response; DOR = duration of response; EFS = event-free survival; FL = follicular lymphoma; IRC = Independent Review Committee; ITT = intent-to-treat; MZL = marginal zone lymphoma; ND = not done; NE = not estimable; ORR = overall response rate; OS = overall survival; PD = progressive disease; PR = partial response; R^2 = lenalidomide plus rituximab; R-mono = rituximab plus placebo; RTNLT = response rate to next anti-lymphoma treatment; SD = stable disease; TTNLT = time to next anti-lymphoma treatment; TTNCT = time to next anti-lymphoma chemotherapy treatment.
Notes: [a]) median estimate is from Kaplan–Meier analysis; [b]) from Cox proportional hazard model adjusting for the three stratification factors: previous rituximab treatment (yes; no), time since last anti-lymphoma therapy (≤2; >2 year), and disease histology (FL; MZL). [c]) from Cox proportional hazard model; [d]) exact confidence interval for binomial distribution; [e]) from CMH test adjusting for the three stratification factors; [f]) from Fisher-Exact test; [g]) from log-rank test adjusting for the three stratification factors; [h]) from log-rank test

Table A1.2: Summary of results from the AUGMENT trial: ITT population (MZL).

Endpoint	**MZL**	
	R^2 (n=31)	**R-mono (n=32)**
Median OS, months (95% CI)[a]	[redacted]	[redacted]
Hazard ratio (95% CI)	[redacted]	
Median PFS, months (95% CI)[a]	24.9 ([redacted])	25.2 ([redacted])
Hazard ratio (95% CI)	[redacted]	
Best response, n (%)		
ORR (CR+PR)	[redacted]	[redacted]
95% CI[d]	[redacted]	[redacted]
p-value	[redacted]	
CR rate	[redacted]	[redacted]
95% CI[d]	[redacted]	[redacted]
p-value	[redacted]	
PR	[redacted]	[redacted]
SD	[redacted]	[redacted]
PD/ death	[redacted]	[redacted]
No evidence of disease	[redacted]	[redacted]
Unknown/ND/Missing	[redacted]	[redacted]
Median TTNLT, months (95% CI)[a]	[redacted]	[redacted]
TTNLT rate at 2 years, % (95% CI)	[redacted]	[redacted]
Hazard ratio (95% CI)	[redacted]	
p-value	[redacted]	

Endpoint	MZL	
	R^2 (n=31)	**R-mono (n=32)**
Median EFS, months (95% CI)[a]	[redacted]	[redacted]
Hazard ratio (95% CI)	[redacted]	
p-value	[redacted]	
Median TTNCT, months (95% CI)[a]	NR	NR
TTNCT rate at 2 years, % (95% CI)	NR	NR
Hazard ratio (95% CI)	NR	
p-value	NR	
RTNLT		
ORR, n (% [95% CI][d])	NR	NR
p-value	NR	
CR, n (% [95% CI][d])	NR	NR
p-value	NR	
DCRR, n/N (%)	[redacted]	[redacted]
95% CI[d]	[redacted]	[redacted]
p-value	[redacted]	
N, Median DOR, months (95% CI)[a]	[redacted]	[redacted]
Hazard ratio (95% CI)[c]	[redacted]	
p-value[e]	[redacted]	
N, Median DOCR, months (95% CI)[a]	[redacted]	[redacted]
Hazard ratio (95% CI)[h]	[redacted]	
p-value	[redacted]	

Source: Response to CL, Table 2, pages 20 and 21.
CI = confidence interval; CR = complete response; DCRR = durable complete response rate, DOCR = duration of complete response; DOR = duration of response; EFS = event-free survival; FL = follicular lymphoma; IRC = Independent Review Committee; ITT = intent-to-treat; MZL = marginal zone lymphoma; ND = not done; NE = not estimable; ORR = overall response rate; OS = overall survival; PD = progressive disease; PR = partial response; R^2 = lenalidomide plus rituximab; R-mono = rituximab plus placebo; RTNLT = response rate to next anti-lymphoma treatment; SD = stable disease; TTNLT = time to next anti-lymphoma treatment; TTNCT = time to next anti-lymphoma chemotherapy treatment.
Notes: [a]) median estimate is from Kaplan–Meier analysis; [b]) from Cox proportional hazard model adjusting for the three stratification factors: previous rituximab treatment (yes; no), time since last anti-lymphoma therapy (≤2; >2 year), and disease histology (FL; MZL). [c]) from Cox proportional hazard model; [d]) exact confidence interval for binomial distribution; [e]) from CMH test adjusting for the three stratification factors; [f]) from Fisher-Exact test; [g]) from log-rank test adjusting for the three stratification factors; [h]) from log-rank test

Summaries of the treatment-emergent adverse event (TEAEs) during AUGMENT for the FL and MZL populations separately are is presented in Tables A1.3 and A1.4, respectively.

Table A1.3: Summary of treatment-emergent adverse events in AUGMENT: FL Safety population

	FL	
	R^2 (n=146)	**R-mono (n=148)**
Number of patients (%)		
Any TEAE	[redacted]	[redacted]

	FL	
	R^2 (n=146)	**R-mono (n=148)**
Len/Pbo related	[redacted]	[redacted]
R related	[redacted]	[redacted]
Grade 3–4 TEAE	[redacted]	[redacted]
Len/Pbo related	[redacted]	[redacted]
R related	[redacted]	[redacted]
Grade 5 TEAE	[redacted]	[redacted]
Any SAE	[redacted]	[redacted]
Len/Pbo related	[redacted]	[redacted]
R related	[redacted]	[redacted]
Any TEAE leading to dose reduction of Len/Pbo	[redacted]	[redacted]
Any TEAE leading to dose interruption of Len/Pbo	[redacted]	[redacted]
Any TEAE leading to dose interruption of R	[redacted]	[redacted]
Any TEAE leading to discontinuation of Len/Pbo	[redacted]	[redacted]
Any TEAE leading to discontinuation of R	[redacted]	[redacted]

Source: Clarification Letter, Table 6, page 21.
FL = follicular lymphoma; Len = lenalidomide; Pbo = placebo; R = rituximab; R^2 = lenalidomide + rituximab; R mono= placebo, rituximab + placebo; SAE = serious adverse event; TEAE = treatment-emergent adverse event.

Table A1.4: Summary of treatment-emergent adverse events in AUGMENT: MZL Safety population

	MZL	
	R^2 (n=30)	**R-mono (n=32)**
Number of patients (%)		
Any TEAE	[redacted]	[redacted]
Len/Pbo related	[redacted]	[redacted]
R related	[redacted]	[redacted]
Grade 3–4 TEAE	[redacted]	[redacted]
Len/Pbo related	[redacted]	[redacted]
R related	[redacted]	[redacted]
Grade 5 TEAE	[redacted]	[redacted]
Any SAE	[redacted]	[redacted]
Len/Pbo related	[redacted]	[redacted]
R related	[redacted]	[redacted]
Any TEAE leading to dose reduction of Len/Pbo	[redacted]	[redacted]
Any TEAE leading to dose interruption of Len/Pbo	[redacted]	[redacted]
Any TEAE leading to dose interruption of R	[redacted]	[redacted]
Any TEAE leading to discontinuation of Len/Pbo	[redacted]	[redacted]
Any TEAE leading to discontinuation of R	[redacted]	[redacted]

Source: Clarification Letter, Table 6, page 21.

Len = lenalidomide; MZL = marginal zone lymphoma; Pbo = placebo; R = rituximab; R2 = lenalidomide + rituximab; R mono = placebo, rituximab + placebo; SAE = serious adverse event; TEAE = treatment-emergent adverse event.

The most common TEAEs, occurring in more than 10% of patients, are presented in Tables A1.5 and A1.6 for FL and MZL patients, respectively.

Table A1.5: Most common treatment-emergent adverse events reported in ≥10% of patients in either treatment arm in AUGMENT: FL Safety population

	FL	
	R^2 (n=146)	**R-mono (n=148)**
Number of patients (%)		
Blood and lymphatic system disorders	[redacted]	[redacted]
Neutropenia	[redacted]	[redacted]
Leukopenia	[redacted]	[redacted]
Anaemia	[redacted]	[redacted]
Thrombocytopenia	[redacted]	[redacted]
Gastrointestinal disorders	[redacted]	[redacted]
Diarrhoea	[redacted]	[redacted]
Constipation	[redacted]	[redacted]
Abdominal pain	[redacted]	[redacted]
Nausea	[redacted]	[redacted]
Infections and infestations	[redacted]	[redacted]
URTI	[redacted]	[redacted]
Nasopharyngitis	[redacted]	[redacted]
General disorders and administration site conditions	[redacted]	[redacted]
Fatigue	[redacted]	[redacted]
Pyrexia	[redacted]	[redacted]
Asthenia	[redacted]	[redacted]
Oedema peripheral	[redacted]	[redacted]
Skin and subcutaneous tissue disorders	[redacted]	[redacted]
Pruritus	[redacted]	[redacted]
Rash	[redacted]	[redacted]
Musculoskeletal and connective tissue disorders	[redacted]	[redacted]
Muscle spasms	[redacted]	[redacted]
Back pain	[redacted]	[redacted]
Respiratory, thoracic and mediastinal disorders	[redacted]	[redacted]
Cough	[redacted]	[redacted]
Dyspnoea	[redacted]	[redacted]
Investigations	[redacted]	[redacted]
Alanine aminotransferase increased	[redacted]	[redacted]
Metabolism and nutrition disorders	[redacted]	[redacted]

	FL	
	R² (n=146)	**R-mono (n=148)**
Decreased appetite	[redacted]	[redacted]
Nervous system disorders	[redacted]	[redacted]
Headache	[redacted]	[redacted]
Injury, poisoning and procedural complications	[redacted]	[redacted]
Infusion related reaction	[redacted]	[redacted]
Eye disorders	[redacted]	[redacted]
Neoplasms benign, malignant and unspecified (including cysts and polyps)	[redacted]	[redacted]
Tumour flare	[redacted]	[redacted]
Psychiatric disorders	[redacted]	[redacted]
Cardiac disorders	[redacted]	[redacted]
Vascular disorders	[redacted]	[redacted]
Source: Clarification Letter, Table 7, pages 22-23. FL = follicular lymphoma; R² = lenalidomide + rituximab; R = placebo, rituximab + placebo; URTI = upper respiratory tract infection.		

Table A1.6: Most common treatment-emergent adverse events reported in ≥10% of patients in either treatment arm in AUGMENT: MZL Safety population

	MZL	
	R² (n=30)	**R-mono (n=32)**
Number of patients (%)		
Blood and lymphatic system disorders	[redacted]	[redacted]
Neutropenia	[redacted]	[redacted]
Leukopenia	[redacted]	[redacted]
Anaemia	[redacted]	[redacted]
Thrombocytopenia	[redacted]	[redacted]
Gastrointestinal disorders	[redacted]	[redacted]
Diarrhoea	[redacted]	[redacted]
Constipation	[redacted]	[redacted]
Abdominal pain	[redacted]	[redacted]
Nausea	[redacted]	[redacted]
Infections and infestations	[redacted]	[redacted]
URTI	[redacted]	[redacted]
Nasopharyngitis	[redacted]	[redacted]
General disorders and administration site conditions	[redacted]	[redacted]
Fatigue	[redacted]	[redacted]
Pyrexia	[redacted]	[redacted]
Asthenia	[redacted]	[redacted]
Oedema peripheral	[redacted]	[redacted]
Skin and subcutaneous tissue disorders	[redacted]	[redacted]

	MZL	
	R^2 (n=30)	**R-mono (n=32)**
Pruritus	[redacted]	[redacted]
Rash	[redacted]	[redacted]
Musculoskeletal and connective tissue disorders	[redacted]	[redacted]
Muscle spasms	[redacted]	[redacted]
Back pain	[redacted]	[redacted]
Respiratory, thoracic and mediastinal disorders	[redacted]	[redacted]
Cough	[redacted]	[redacted]
Dyspnoea	[redacted]	[redacted]
Investigations	[redacted]	[redacted]
Alanine aminotransferase increased	[redacted]	[redacted]
Metabolism and nutrition disorders	[redacted]	[redacted]
Decreased appetite	[redacted]	[redacted]
Nervous system disorders	[redacted]	[redacted]
Headache	[redacted]	[redacted]
Injury, poisoning and procedural complications	[redacted]	[redacted]
Infusion related reaction	[redacted]	[redacted]
Eye disorders	[redacted]	[redacted]
Neoplasms benign, malignant and unspecified (including cysts and polyps)	[redacted]	[redacted]
Tumour flare	[redacted]	[redacted]
Psychiatric disorders	[redacted]	[redacted]
Cardiac disorders	[redacted]	[redacted]
Vascular disorders	[redacted]	[redacted]

Source: Clarification Letter, Table 7, pages 22-23.
MZL = marginal zone lymphoma; R^2 = lenalidomide + rituximab; R = placebo, rituximab + placebo; URTI = upper respiratory tract infection.

Appendix 2: MAIC reporting checklist

According to the NICE DSU, the following themes should be considered and addressed explicitly, when reporting population-adjusted analyses (See TSD 18, pages 64-65).[41]

Criteria	Addressed in CS (Y/N)	ERG Comments
1. The variables available in each study should be listed, along with their distributions (e.g. through box plots or histograms).	Y	The variables were listed along with summary statistics, although there were no plots of their distributions
2. Sufficient covariate overlap between the populations should be assessed: - for population reweighting methods (such as MAIC), the number of individuals assigned zero weight should be reported; - for outcome regression methods (such as STC), the amount of extrapolation required should be considered. - For anchored comparisons this applies only to effect modifiers (see point 2); - for unanchored comparisons all variables relevant to outcome should be presented.	Y	The CS used a MAIC and the details of the weighting, number of zero weights were provided.
3. Evidence for effect modifier status should be given, along with the proposed size of the interaction effect and the imbalance between the study populations.	N	No information about those variables considered to be effect modifiers and their interaction with the treatment effect.
4. The resulting potential bias reduction compared with a standard indirect comparison may be calculated by multiplying the interaction coefficient by the difference in means.	N	For some analyses there was also an unadjusted indirect comparison but there was no estimate of the bias reduction.
5. The distribution of weights should be presented for population weighting analyses, and used to highlight any issues with extreme or highly variable weights.	Y	Histograms showing the distribution of the weights were provided
6. Presentation of the effective sample size may also be useful.	Y	The ESS was reported for each matched analysis
7. ESS may be approximated using equation (7) – which is likely to be an underestimate – but provides clear warning where inferences are being made based on just a small number of individuals Measures of uncertainty, such as confidence intervals, should always be presented alongside any estimates. - Care should be taken that uncertainty is appropriately propagated through to the final estimates. - For outcome regression methods, uncertainty is fully propagated for predictions into the	Y	Confidence intervals were reported. Standard errors for the survival analyses (Cox and parametric models) were calculated using robust sandwich estimators. A further sensitivity analysis was performed which estimated standard errors using bootstrapping.

Criteria	Addressed in CS (Y/N)	ERG Comments
aggregate population by the outcome regression model. - For population reweighting methods, a robust sandwich estimator (as typical for MAIC) provides estimates of standard error which account for all sources of uncertainty. - Other techniques include bootstrapping and Bayesian methods.		
8. For an unanchored comparison, estimates of systematic error before and after population adjustment should be presented	N	No information
9. Present estimates for the appropriate target population using the shared effect modifier assumption if appropriate, or comment on the representativeness of the aggregate population to the true target population.	Y	It was reported that the MAIC results are only application to the population of the specific comparator trials (the trial providing the summary characteristics). Not all relevant covariates could be included in all analyses so "*the key assumption of the MAIC may not hold, and the results should be interpreted cautiously*"
10. In order to convey some clarity about the impact of any population adjustment, the standard indirect comparison estimate should be presented alongside the population-adjusted indirect comparison if an anchored comparison is formed; for an unanchored comparison, a crude unadjusted difference should be presented alongside the MAIC/STC estimate.	N	The statistical report stated that unanchored indirect comparisons were performed but the results were not reported, or presented in the company submission alongside the MAIC results.