Venetoclax in combination with rituximab for treating relapsed or refractory chronic lymphocytic leukaemia [ID1097]

Produced by: Warwick Evidence

Authors: Chidozie Nduka, Research Fellow, Warwick Evidence
Felix Achana, Senior Research Fellow, Warwick Evidence
Daniel Gallacher, Research Associate, Warwick Evidence
Xavier Armoiry, Senior Research Fellow, Warwick Evidence
Rachel Court, Information Specialist, Warwick Evidence
Kate Evans, Research Project Administrator, Warwick Evidence
Renata Walewska, Consultant Haematologist, Royal Bournemouth Hospital
Hema Mistry, Assistant Professor, Warwick Evidence

Correspondence to: Dr Hema Mistry
Warwick Evidence
Warwick Medical School
University of Warwick
Coventry, CV4 7AL, UK
Tel: +44 (0)2476 151183
Email: Hema.Mistry@warwick.ac.uk

Date completed: 29th August 2018

Source of funding: This report was commissioned by the NIHR HTA Programme as project number 17/156/04

Declared competing interests of the authors

None

Acknowledgements

Professor James Mason provided overview of the topic area. Professor Norman Waugh read the report and provided advice.

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. Copyright belongs to The University of Warwick.

This report should be referenced as follows

Venetoclax in combination with rituximab for treating relapsed or refractory chronic lymphocytic leukaemia: A Single Technology Appraisal. Warwick Evidence, 2018.

Contributions of authors

Chidozie Nduka (Research Fellow) helped co-ordinate the report, and reviewed and critiqued the clinical effectiveness evidence; Felix Achana (Senior Research Fellow) reviewed and critiqued the indirect and matched adjusted indirect comparisons, and the cost-effectiveness evidence and undertook additional analyses; Daniel Gallacher (Research Associate) reviewed and critiqued the statistical and the survival analysis and undertook additional analyses; Xavier Armoiry (Senior Research Fellow) reviewed and critiqued the clinical effectiveness evidence and the indirect and matched adjusted indirect comparisons; Rachel Court (Information Specialist) critiqued the company searches and undertook additional searches; Kate Evans (Research Project Administrator) reviewed and critiqued the background section; Renata Walewska (Consultant Haematologist) provided expert clinical advice; Hema Mistry (Assistant Professor) co-ordinated the project and the report, and reviewed and critiqued the cost-effectiveness evidence.

Word count: 33,220 (including references and appendices)

Please note that: Sections highlighted in yellow and underlined are [redacted] Sections highlighted in aqua and underlined are 'commercial in confidence' (CIC). Figures that are CIC have been bordered with blue.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

DEFINITION OF TERMS AND LIST OF ABBREVIATIONS

Del(17p)	Deletion of the short arm of chromosome 17
AIC	Akaike information criteria
ALC	Absolute lymphocyte count
AEs	Adverse events
ALT	Alanine aminotransferase
AST	Aspartate Transaminase
BCRi	B-cell receptor inhibitor
BCSH	British Committee for Standards in Haematology
BR	Bendamustine plus rituximab
BSC	Best supportive care
BSA	Body surface area
CEAC	Cost-effectiveness acceptability curves
CI	Confidence interval
CLL	Chronic lymphocytic leukaemia
CIRS	Cumulative Illness Rating Scale
CIT	Chemo-immunotherapy treatment
CRCL	Creatinine Clearance
CR	Complete response rate
CRi	CR with incomplete hematologic recovery
CRD	Centre for Reviews and Dissemination
CS	Company submission
CSR	Clinical study report
CT	Computerised tomography
CTCAE	Common terminology criteria for adverse events
ECOG	Eastern Cooperative Oncology Group
EFS	Event-free survival
EMA	European Medicines Agency

ERG	Evidence Review Group
EQ-5D-3L	EuroQoL five-dimension 3-level version
FCR	Fludarabine, cyclophosphamide, rituximab
HMRN	Haematological Malignancy Research Network
HR	Hazard ratio
HRQoL	Health-related quality of life
HTA	Health Technology Assessment
IC	Indirect comparison
ICER	Incremental cost-effectiveness ratio
IDELA+R	Idelalisib in combination with rituximab
IGHV	Immunoglobulin heavy-chain variable
IPD	Individual patient data
IRC	Independent review committee
IV	Intravenous
iwCLL	International Workshop on Chronic Lymphocytic Leukaemia
KM	Kaplan-Meier
LCI	Lower confidence interval
LDH	Lactate dehydrogenase
LY	Life years
MAIC	Matched adjusted indirect comparison
MRD	Minimal residue disease
NCI	National Cancer Institute
NMA	Network meta-analysis
NHS	National Health Service
NICE	National Institute of Health and Care Excellence
ORR	Overall response rate
OS	Overall survival
OWSA	One-way sensitivity analysis
PFS	Progression free survival
PH	Proportional hazards
PICOS	Population, intervention, comparator, outcome
PPS	Post-progression survival

PR	Partial response/remission
PSS	Personal social services
QALY	Quality-adjusted life year
RCT	Randomised controlled trial
R/R	Relapsed or refractory
SAEs	Serious adverse events (SAEs)
SC	Subcutaneous
SLL	Small lymphocytic lymphoma
SUCRA	Surface under the cumulative ranking curve
TA	Technology appraisal
TLS	Tumour Lysis Syndrome
ToT	Time to treatment
TP53	Mutation in the TP53 gene
UCI	Upper confidence interval
VEN+R	Venetoclax in combination with rituximab

1 SUMMARY

Chronic lymphocytic leukaemia (CLL) is a cancer that starts in the blood stem cells. Stem cells are basic cells that develop into different types of cells that have different functions. As the stem cells of the blood develop, they become blast cells, which are immature blood cells. In leukaemia, there is an overproduction of blast cells. These blast cells do not develop into mature blood cells. Over time, the blast cells crowd out normal blood cells so that these normal cells are unable to perform their functions. When leukaemia is diagnosed, these blast cells may be called leukaemia cells. In lymphocytic leukaemias, these leukaemia cells develop from abnormal lymphoid stem cells.

Treatments for CLL include: watchful waiting, chemotherapy, targeted therapy, surgery, stem cell transplant, and supportive therapy. The type(s) of treatment offered is based on a number of factors including: stage, age, overall health, and personal preferences. The objective of the final scope to appraise the clinical and cost-effectiveness of targeted therapy (venetoclax in combination with rituximab) within its marketing authorisation for treating relapsed or refractory chronic lymphocytic leukaemia.

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

The company specifies that patients with relapsed or refractory (R/R) CLL were eligible to be included as part of the submission only if they previously received chemo-immunotherapy (CIT). While the final scope also describes patients with R/R CLL as the target population for the technology appraisal, the ERG clinical advisor considers that CLL patients with deletion of the short arm of chromosome 17 (del(17p)) / mutation in the TP53 gene (TP53 mutation) may never receive CIT, given that these patients receive ibrutinib as first-line in clinical practice. The intervention in the submission is venetoclax in combination with rituximab (VEN+R), which is the same as the final scope. Venetoclax is given until disease progression or unacceptable toxicity, or for a maximum duration of two years, whichever occurs first. While the two-year stopping rule seems arbitrary and not based on any empirical evidence comparing different stopping rules (e.g. 18 months versus 24 months), the ERG's clinical advisor agrees with the two-year stopping rule of venetoclax as it is anticipated that most patients would have achieved negative minimal residual disease (MRD) status by this time, otherwise a different line of therapy

must be considered. Single-agent ibrutinib or idelalisib-rituximab combination (IDELA+R) were the main comparators presented in the decision problem and final scope, with ibrutinib considered more clinically relevant by the ERG's clinical advisor: ibrutinib is more effective and less toxic compared to IDELA+R. The outcomes of interest (progression-free survival, overall survival, response rates, minimal residual disease status, adverse events, and health-related quality of life) were also clinically relevant and consistent with the final scope and trial evidence submitted (MURANO, RESONATE, and Study 116). Given that data from the key trial evidence (MURANO) was not mature enough to estimate the overall survival (OS), the ERG also agrees that progression free survival (PFS) was a reasonable primary endpoint. However, the ERG maintains that OS is a much more reliable outcome than PFS.

1.2 Summary of clinical effectiveness evidence submitted by the company

The MURANO trial of VEN+R compared against Bendamustine-Rituximab (BR) combination showed that the risk of progression or death (PFS) was reported to be significantly lower in the VEN+R group compared to the BR group after a median follow-up duration of 23.8 months, as assessed by the investigators (hazard ratio (HR), 0.17; 95% confidence interval (CI) 0.11 to 0.25) and by an independent review committee (HR 0.19; 95% CI 0.13 to 0.28). VEN+R was also superior to BR in terms of OS (HR 0.48; 95% CI 0.25 to 0.90) and MRD clearance rates in blood (absolute difference at any time during the trial 60.4%; 95% CI 52.3% to 68.6%) and bone marrow (absolute difference at any time during the trial 25.8%; 95% CI 19.0% to 32.6%). In the absence of a head-to-head trial comparing VEN+R to ibrutinib (or IDELA+R) and RCT evidence providing a comparator common to VEN+R and ibrutinib (or IDELA+R), the company also identified from the literature search other trials (RESONATE and Study 116) suitable for performing an unanchored matched adjusted indirect treatment comparison (MAIC). Respectively, the RESONATE and Study 116 trials showed ibrutinib and IDELA+R to be significantly more effective than their comparators (ofatumumab and rituximab-placebo).

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

The ERG had no major concerns over the statistical methods in MURANO, RESONATE and Study 116 trials. The ERG acknowledges that patients in MURANO who would not have been eligible for these comparator trials (RESONATE or Study 116) were appropriately excluded from

the MAIC. To adjust for any residual cross-trial differences in the MAIC, patients in the MURANO trial were weighted such that their weighted mean baseline characteristics matched those reported for the RESONATE and Study 116 trials.

The ERG reviewed the results of the unanchored MAIC with emphasis on the comparison of VEN+R with ibrutinib. These results showed that VEN+R has a [redacted] progression or death (PFS) events, however, this treatment effect remains statistically not significant given the wide confidence intervals in the hazard ratio ([redacted] For OS, the MAIC results showed that VEN+R lowered the rate of death events overall by [redacted] compared to ibrutinib ([redacted] Given this degree of contrast between the PFS and OS benefits, the ERG considers that the magnitude of the latter may not be realistic. Sensitivity analysis undertaken by the ERG indirectly comparing estimates of the treatment effect of VEN+R (from the MURANO trial) against single-agent ibrutinib (from a previously published indirect treatment comparison of ibrutinib vs BR vs ibrutinib+BR) validated the ERG's concerns: HRs = [redacted] and [redacted] for PFS and OS respectively.

1.4 Summary of cost-effectiveness submitted evidence by the company

The company conducted a systematic literature search to identify published cost-effectiveness studies and economic models, but found none comparing the cost-effectiveness VEN+R with ibrutinib or IDELA+R as treatment options for R/R CLL. Thus, the company developed a *de novo* partitioned survival model (consistent with the NICE reference case) to simulate lifetime economic costs and outcomes associated with the comparator interventions from the UK NHS and personal social services (PSS) perspective. The base-case model simulated survival outcomes for patients on VEN+R based on evidence from the MURANO trial with extrapolation over a lifetime horizon. In the model, this was assumed to be 30-years for an R/R CLL cohort with a mean age of 64 years. Survival outcomes for comparator interventions were generated by applying hazard ratios derived from unanchored MAIC comparisons to model predictions of outcomes for patients on VEN+R. The CS base-case applied a discount rate of 3.5% per annum to both costs and outcomes over the modelled time-horizon. The model suggested that VEN+R dominated ibrutinib (i.e. VEN+R was cheaper and generated more quality-adjusted life years

(QALYs) compared with ibrutinib). For the comparison with IDELA+R, the model generated an incremental cost-effectiveness ratio (ICER) of [redacted] per QALY gained for VEN+R. Based on list price comparisons, probabilistic sensitivity analysis suggested that VEN+R was close to [redacted] probability of being cost-effective at £20,000 per QALY compared to ibrutinib and over [redacted] probability of being cost-effective at £20,000 per QALY compared to IDELA+R. Sensitivity analyses suggest the ICER was mainly sensitive to the hazard ratio for overall survival, the modelled time horizon and the methods used to extrapolate survival outcomes over longer time horizon.

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

The ERG found the company's approach to economic modelling appropriate and consistent with NICE reference case. The model structure is similar to economic models that informed two previous appraisals in CLL (TA359 and TA587). The ERG is satisfied with the approach used to estimate health-state utilities and adverse events disutilities. Costs relevant to the decision problem appears to have been appropriately accounted for in the model, although a minor error in calculation of intervention costs had meant that rituximab costs were included during (rather than after) the dose escalation stage of the VEN+R treatment regimen. As stated above, for the comparison with ibrutinib, the ERG had major reservations about robustness of the company's MAIC analyses, and believes any uncertainty in the hazard ratio would translate into uncertainties in cost-effectiveness that would be difficult to quantify. For the comparison with IDELA+R, the ERG does not believe evidence was presented to estimate efficacy of VEN+R vs. IDELA+R with a degree of confidence. Overall, the key drivers of cost-effectiveness were the OS hazard ratio, the methods used to extrapolate survival outcomes and the 2-year fixed treatment duration which considerably lowered treatment costs for VEN+R. The ERG believes these parameters are highly uncertain, the former because of the uncertainty emanating from the MAIC analysis mentioned above and the latter two, because the immaturity of the MURANO data meant no robust data is currently available to validate the 2-year fixed treatment duration.

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

1.6.1 Strengths

The key strength of the company's submission relies on the appropriateness and good methodological quality of the trials included in the MAIC.

The ERG also confirms that no eligible study was missing from the MAIC.

The structure of the economic model is similar to economic models used in previous NICE technology appraisals of interventions in CLL. Health-state utility values were taken from previous NICE appraisal committees' most preferred base-case model in CLL (TA487) and a similar approach to estimation of disutility associated with adverse events was applied. Extensive sensitivity analyses suggests results were mostly robust to alternative parameter inputs and model assumptions considered in the CS.

1.6.2 Weaknesses and areas of uncertainty

The absence of head-to-head trials comparing VEN+R against single-agent ibrutinib is perhaps the most obvious weakness in the company's submission.

The ERG also considers that the immaturity of OS data from the MURANO trial is a major weakness in the company's submission as it contributes significantly to the implausible OS results in the MAIC, with OS reduced by [redacted] and PFS by [redacted]

The wide confidence intervals for the primary endpoint (PFS) HRs suggest that the treatment effect of VEN+R may be somewhat biased.

The ERG has major reservations about robustness of the companies MAIC analyses, and believes any uncertainty in the hazard ratio would translate into uncertainties in the cost-effectiveness analyses that are difficult to quantify.

1.7 Summary of exploratory and sensitivity analyses undertaken by the ERG

The ERG focused its exploratory analyses around HRs for PFS and OS for VEN+R vs. ibrutinib or IDELA+R and the methods used to extrapolate survival over a lifetime horizon. These are the main components of the company's economic model where ERG believed the evidence base was weakest, and the ERG identified these as the key drivers of cost-effectiveness in the CS sensitivity analyses. For the comparison with ibrutinib, the ERGs preferred a base-case model that used HRs generated from indirect comparison analysis, and joint-Gamma model to extrapolate survival outcomes, to suggest VEN+R was considerably cheaper (incremental costs of [redacted]) but also generated fewer QALYs (incremental QALYs of -0.39) compared with ibrutinib with an ICER of [redacted] per QALY lost based on list price comparisons. The ERG's exploratory base-case analyses were not conducted for the comparison with IDELA+R due to lack of robust evidence on the relative effectiveness of the two interventions.

2 BACKGROUND

2.1 Critique of company's description of underlying health problem

The company submission (CS) provides an overview of chronic lymphocytic leukaemia (CLL) (CS section B 1.3.1). The CS correctly states that 'CLL is the most common of the chronic leukaemias'.[1] The CS describes CLL as a disease of unknown aetiology characterised by the accumulation of mature B lymphocytes in blood, lymph nodes, spleen, liver, and bone marrow. This description is broadly consistent with the final scope provided by the National Institute of Health and Care Excellence (NICE). According to the CS, this accumulation of B lymphocytes can lead to a wide variety of symptoms that manifest as fatigue, loss of appetite, weight loss, night sweats and shortness of breath on exertion. However, it should be noted that CLL is often asymptomatic and diagnosed by chance. The clinical pattern ranges from no treatment needed to rapid progression. These symptoms are also consistent with those described by the British Committee for Standards in Haematology (BCSH).[2] The CS identifies recurrent genetic abnormalities (deletions or mutations) as the main cause of CLL. The disease is subject to clonal variation during the disease course (due to mutation of the tumour suppressor gene TP53) which mediates resistance to chemotherapy. TP53 dysregulation is observed in 5-10% of untreated CLL patients and present in 40-50% of patients with refractory disease. The ERG finds research to support these statements.[3]

There were 3,709 new diagnoses of CLL in 2015 which is slightly higher than reported in the CS.[4] The ERG agrees that the age-standardised incidence of CLL is 6.5 per 100,000.[4] Based on a study by Shanafelt et al (2010), the company states that survival of CLL patients is observed to be significantly shorter than that of the age-matched general population ($p < 0.001$).[5] However, this study was conducted in Minnesota, USA. The company does not provide incidence statistics by age or survival rates. According to Cancer Research UK, CLL incidence is strongly related to age, with the highest incidence rates being in older people. In the UK in 2013-2015, on average each year more than 4 in 10 (43%) of new cases were in people aged 75 and over'.[6] More so, the five-year survival rate for men in the UK is 51% - 72% and 73% - 81% for women.[7]

The company provides an overview of the disease burden (CS section B.1.3.2) for symptomatic CLL patients. They discuss reduction of health-related quality of life (HRQoL) and attribute it primarily to disease progression and fatigue, which the ERG verifies to be accurate.

2.2 Critique of company's overview of current service provision

The current treatment of CLL is outlined in section B.1.3.4 and is consistent with the final scope. The CS makes reference to NICE guidance and guidelines published by the BCSH. Key recommendations are summarised in CS Table 3 and pathways shown in Table 3 and Figure 1. The current treatment pathway depends on diagnosis and previous treatments. Venetoclax monotherapy is recommended by NICE technology appraisal (TA) guidance TA487 as a second line treatment for patients with del(17p) and/or TP53 mutation experiencing disease progression after receiving B-cell receptor inhibitor (BCRi) treatment.[8] NICE TA429 recommends ibrutinib for patients who have had at least 1 prior chemo-immunotherapy treatment (CIT).[9] This is in alignment with the final scope. Additionally, NICE TA359 recommends idelalisib in combination with rituximab for adults with relapsed or refractory (R/R) CLL disease.[10] However, the CS states that ibrutinib is the more commonly used BCRi therapy due to toxicity concerns associated with idelalisib and ibrutinib being more effective than idelalisib in combination with rituximab (IDELA+R). The ERG clinical advisor agrees that this treatment strategy reflects the current position of the National Health Service (NHS).

Unmet need

The CS considers the high unmet need for the treatment of CLL patients with relapsed or refractory disease and high risk genetic subtypes (including TP53 dysregulation). They describe a need to identify effective therapies with alternative mechanisms of action and acceptable side effect profiles (CS section B.1.3.1). The CS states that early intervention with chemotherapy does not improve the natural history of the disease, may drive clonal evolution and later treatment resistance and hence, therapy is only recommended for patients with rapidly progressive or symptomatic disease. The company suggests that once treatments are stopped, due to disease progression and no other treatment options available, survival is poor (CS section B.1.3.2). The company also details that there is increased negative impact on both the patients' and their carers' HRQoL as the disease progresses. They highlight an increased economic burden reporting that

R/R CLL patients have the highest resource use among CLL patients (CS section B.1.3.2),), which the ERG clinical advisor suggests is plausible.

Furthermore according to the CS (section B.1.3.5) patients post CIT with deletion of the short arm of chromosome 17 (del(17p)) / mutation in the TP53 gene (TP53) have fewer treatment options than non-del(17p)/TP53 patients. BCRi therapies (e.g. ibrutinib) are highly effective in this subgroup, but are associated with an indefinite treatment period and do not result in high rates of undetectable minimal residue disease (MRD). Therefore, the CS finds there is an unmet need for therapies demonstrating improved survival outcomes in both del(17p)/TP53 and non-del(17p)/TP53 sub-populations and that demonstrate potential to achieve MRD-negative status.

Treatment pathway of VEN+R

The company anticipates venetoclax in combination with rituximab (VEN+R) is likely to be used for patients with CLL who have received at least one prior therapy (CS section B.1.3.5, figure 1) within the UK NHS, specifically post-CIT. However, the ERG clinical advisor disagrees with the positioning of VEN+R in the treatment pathway for patients with del(17p) and/or TP53 mutation because CIT is generally not considered a treatment option in these patients.

3 Critique of company's definition of decision problem

The company described the decision problem in Table 1 of the submission (CS, pg 15-17).

3.1 Population

In their decision problem, the company describes adults with R/R CLL as the target population for the technology appraisal, which is broadly consistent with the final scope and the trial populations in the key evidence submitted.[11-13]

Following consultations with their clinical experts, the company further specify that patients were eligible to be included as part of the submission only if they previously received chemo-immunotherapy – in line with the anticipated position of the technology (VEN+R) in the treatment pathway for R/R CLL in the UK (CS, Figure 1). However, the ERG is concerned that restricting the target population to patients post CIT potentially excludes CLL patients with del(17p) and/or TP53 mutation. In this high-risk subgroup, the ERG clinical advisor questions the position of VEN+R as illustrated in the proposed treatment pathway in the CS (CS Figure 1). The ERG clinical advisor considers that patients with del(17p)/TP53 mutation CLL may never receive CIT, given that these patients receive BCRi therapy (ibrutinib) as first-line in clinical practice.

Although the company recognises ibrutinib as the mainstay for the first-line treatment of del(17p)/TP53 mutation CLL as recommended in NICE TA429, they maintain that a small number of these patients receive CIT as first-line treatment (CS pg 26). The ERG considers this evidence to be largely anecdotal, and should not have informed the population selection in the decision problem.

3.2 Intervention

The intervention in the submission is venetoclax in combination with rituximab, which is the same as the final scope. The company provides a description of the technology and the mechanism of action of venetoclax (CS Table 2, pg 18) which the ERG's clinical advisor confirms to be accurate. According to the summary of product characteristics, VEN+R is indicated for the treatment of adult patients with CLL who have received at least one prior

therapy. Venetoclax is initially administered (orally) in weekly dose increments up to 400 mg at week 5. At this time, rituximab is commenced simultaneously as a monthly injection up to a total of six months/cycles (375 mg/m^2 in the first cycle and 500 mg/m^2 in cycles 2 to 6). From week 5 onwards, venetoclax is given at a dose of 400 mg daily up to a maximum of two years. The ERG clinical advisor agrees with this two-year stopping rule, irrespective of the treatment outcome, as time limited treatment would increase compliance, would be a more acceptable option to some patients and reduce the cost of the treatment. However, it is anticipated that most patients would have achieved negative MRD status by this time.

3.3 Comparators

Ibrutinib and IDELA+R were listed comparators in the decision problem and final scope. The CS stated that in the absence of head-to-head trials comparing VEN+R with ibrutinib or IDELA+R, together with the absence of randomised controlled trial (RCT) evidence that could have enabled an indirect treatment comparison using network meta-analysis, the company carried out a matched adjusted indirect comparison (MAIC) of VEN+R versus single-agent ibrutinib.

In contrast to the final scope, the company deemed best supportive care (BSC) inappropriate as a comparator in the appraisal, while asserting that BSC is only reserved for later lines of therapy after all treatment options have failed. The ERG clinical advisor agrees that BSC is the last course of action given for palliation as opposed to disease modification.

Although venetoclax monotherapy was not included in the NICE scope and therefore was not discussed by the company, the ERG's clinical advisor has emphasized that venetoclax monotherapy appears to have a more favourable safety profile compared to ibrutinib, and is the mainstay of treatment in CLL patients who do not tolerate ibrutinib irrespective of TP53 mutation status.

3.4 Outcomes

The outcomes of interest in the final scope match those specified in the decision problem as well as trial evidence submitted.

The ERG has noted that its clinical advisor considered MRD to be the single most important clinical indicator to assess in trials in patients with CLL, emphasising strongly that a MRD negative status is the closest a patient gets to a cure. However, the company did not provide MAIC analyses of the MRD status, when the ERG requested this at the clarification stage.

The ERG also agrees that progression free survival (PFS) was a reasonable primary endpoint considering that data from the MURANO trial was not mature enough to estimate the overall survival (OS) and that PFS is a valid surrogate outcome for OS.[14]

3.5 Other relevant factors

The CS reports that there are no equality issues presented by VEN+R. The company also anticipates that the European Medicines Agency (EMA) license for VEN+R will be issued in [redacted].

4 CLINICAL EFFECTIVENESS

4.1 Critique of the methods of review(s)

The company undertook a broad systematic review aimed at identifying randomised and non-randomised clinical trials investigating the clinical effectiveness of VEN+R and comparator interventions for treating patients with R/R CLL. Comparator interventions include those defined in the company decision problem for this submission and many others as reported in CS Table 5, pg 29. One trial of VEN+R (MURANO) was identified and considered relevant to the decision problem.[11] Overall the ERG found the company's systematic review to be of reasonable quality. Table 1 summarises the ERG's quality assessment of the company's systematic review.

Table 1: Quality assessment of the CS systematic review of clinical effectiveness

CRD Quality Item	Yes/No/Uncertain with comments
1. Are any inclusion/exclusion criteria reported relating to the primary studies which address the review question?	Yes
2. Is there evidence of a substantial effort to search for all relevant research?	Yes
3. Is the validity of included studies adequately assessed?	The validity of the MURANO trial alone was assessed, including issues pertaining to the external validity of the study outcomes (CS Table 10, pg 39).
4. Is sufficient detail of the individual studies presented?	Sufficient details were presented for the MURANO trial alone
5. Are the primary studies summarised appropriately?	The MURANO trial alone was summarised appropriately.

4.1.1 Searches (Description of company's search strategy)

Although the company did not search trial registers and Health Technology Assessment (HTA) agencies for studies eligible for their systematic review, the ERG considers the literature searches to be comprehensive using a number of relevant bibliographic databases (such as MEDLINE and Embase via the ProQuest interface). The searches ─ undertaken on 21 July 2017 and updated on 30 April 2018 ─ were conducted using appropriate search terms; without any restriction on publication date (except for the 2014 publication date limit applied to the search for conference proceedings); and excluded published letters, notes, errata and editorials. While restricting the searches to studies published in English language may have introduced some language bias, the ERG has found no missing relevant studies published in a different language. The ERG also

reviewed the list of studies excluded from the MAIC and deemed them irrelevant to the company's decision problem and final scope. However, of the 49 studies potentially eligible for the MAIC (CS Figure 1, pg 32), the ERG could only review 48 full-texts provided by the company at the ERG's request (Clarification Response C1). Nonetheless, additional searches undertaken by the ERG identified no missing studies that were relevant to the decision problem.

4.1.2 Inclusion/exclusion criteria used in the study selection

Eligibility criteria for the CS systematic review are summarised in CS Table 5, pg 29. Adults with established R/R CLL were eligible for the company's systematic review, which matches the NICE final scope. However, the inclusion criteria for the target population is broader than the company's decision problem, which specifies that the target population must include patients who have received prior chemo-immunotherapy. The ERG critiqued the rationale for this distinction in section 3.1. The interventions (VEN+R), comparators (ibrutinib and IDELA+R) and study outcomes listed in the final scope and decision problem were also specified as part of the inclusion criteria.

4.1.3 Critique of data extraction

The ERG considers that the company conducted the study selection (two independent reviewers with third reviewer/strategic advisor resolving discrepancies) and data extraction (two independent reviewers with third reviewer/strategic advisor resolving discrepancies) appropriately. However, no information is provided on the method of data extraction.

4.1.4 Quality assessment of key trials

The company provided a quality assessment of its own MURANO trial using the minimum criteria for assessing risk of bias in RCTs as set out in the Centre for Reviews and Dissemination (CRD) guidance for undertaking reviews in health care and the NICE single technology appraisal user guide (CS Table 13, pg 46). In addition, the company addresses issues about the generalisability of the trial findings to clinical practice in England. The ERG conclude that these are sufficient, however, the company has not presented quality assessments of the RESONATE[12]

and Study 116[13] trials in the main submission. Table 2 summarises the ERG's critique of the company's quality appraisal for MURANO.

Although the MURANO trial was open-label using two different routes of administration for VEN+R (oral venetoclax + intravenous rituximab) and one route for BR (intravenous), the ERG considers that this trial need not have been open-label as oral placebos could have been administered in the BR arm.

The company suggests that the MURANO trial is reflective of clinical practice in England because BR was considered the most effective treatment for managing R/R CLL patients with del(17p) at the time the trial was initiated.

Table 2: Quality assessment of the MURANO trial

Question	Company's response	ERG's response	Rationale for ERG's response	ERG's rationale for discrepancy
Was randomisation carried out appropriately?	Yes.	Yes	Participants were randomised 1:1 using a web-based randomisation system	N/A
Was the concealment of treatment allocation adequate?	The MURANO trial was open label, using two different methods of administration (oral or intravenous (IV))	Unclear	Protecting the allocation sequence before and until assignment is not described, but an Interactive Voice/Web Response System is used to randomize patients, which may also serve this purpose.	Allocation concealment was not reported in the submission or MURANO protocol or report
Were the groups similar at the outset of the study in terms of prognostic factors?	Yes	Yes	Baseline characteristics were similar between treatment arms	N/A
Were the care providers, participants and outcome assessors blind to treatment allocation?	No	No	This was an open label trial which suggests that the participants and investigators were not blind to treatment allocation. However, the ERG maintains that the outcome assessors could have been blinded.	N/A
Were there any unexpected imbalances in drop-outs between groups?	No	No	Although there was a significant difference in withdrawal rates between	N/A

			VEN+R and BR (4% vs 10%, $p < 0.02$), the ERG is not surprised about this given the open-label nature of the trial. The ERG would be more concerned if withdrawal rates were much higher in the VEN+R arm compared to BR, especially considering that BR is administered for a total of six 28-day cycles and VEN+R is given for two years.	
Is there any evidence to suggest that the authors measured more outcomes than they reported?	No	No	All efficacy outcomes reported in the results were pre-specified in the protocol	N/A
Did the analysis include an intention-to-treat analysis? If so, was this appropriate and were appropriate methods used to account for missing data?	Yes	Yes	Although seven patients in the BR arm withdrew from the trial just after randomisation, these patients were accounted for in the efficacy analyses	

4.1.5 Evidence Synthesis

In the absence of a head-to-head trial comparing VEN+R to any of the comparators listed in the final scope, the company sought to perform a matched adjusted indirect comparison of VEN+R against these comparators by screening the search records for relevant comparator trials. Two trials (RESONATE and Study 116) were identified for this purpose and deemed relevant to the decision problem.

The ERG considers that the criteria for including studies in the MAIC as stated in CS section B.2.9.3 (pg 43 and 44) are not exhaustive. For instance, while the company states that study outcomes and follow-up duration of survival data had to be similar between MURANO and its comparator trials, it is not stated that trial populations had to be comparable across these trials. Nonetheless, the ERG also considers that the aim of MAIC is to create comparable groups by using the IPD of one to remove people till the remaining group matches the recruits in the other trial.

Although no formal quality appraisal was presented for the MAIC, the ERG considers that inclusion/exclusion criteria were fairly matched across the MURANO, RESONATE and Study 116 trials. For instance, all included patients must have been treated previously for CLL and have an Eastern Cooperative Oncology Group (ECOG) score of 0 or 1. Patients in the MURANO trial who would not have been eligible for the RESONATE and Study 116 trials, were appropriately excluded from the MAIC.[15] Only quantitative effect-modifiers (prior to matching) were selected as baseline matching characteristics for the MAIC, the ERG considers this method of variable selection to be sufficiently rigorous. However, the ERG is unable to determine how the RESONATE and Study 116 trials were assessed for availability of individual patient-level data as implied in the CS (Section B.2.9.4, pg 67).

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

Evidence for the clinical effectiveness on VEN+R comes from a single pivotal RCT. The MURANO trial (ClinicalTrials.gov Identifier: NCT02005471) was a phase III open-label, multicentre, active treatment controlled RCT sponsored by the company. The results are currently being reviewed by the EMA as part of the process aimed to extend marketing authorisation of venetoclax, which is already licensed for treating CLL as a single agent. The trial was designed to investigate the use of venetoclax in combination with rituximab (VEN+R) in patients with R/R CLL.

The dosing schedule for VEN+R is described in section 3. Interestingly, unlike single agent ibrutinib which is licensed for the same indication as VEN+R (patients with R/R CLL), venetoclax is given for a maximum of two years. The comparator in the MURANO trial was bendamustine plus rituximab (BR) where bendamustine was given intravenously (70mg/ m^2 on days one and two of each 28-day cycle) and rituximab was administered as described for VEN+R in section 3.

The MURANO trial was commenced in March 2014 and all participants were randomised by September 2015. The clinical cut-off date was May 2017. The randomisation ratio was 1:1 between treatment arms and stratified according to del(17p) status, responsiveness to previous

therapy, and geographic region. Cross-over to the VEN+R arm in the event of disease progression was not allowed, however, treatment post-progression was at the investigators' discretion. Key inclusion criteria are reported in CS Table 7 (pg 33) including age ≥18 years, CLL with R/R status, no more than three previous treatments, and an ECOG performance status score of 0 or 1. Key exclusion criteria were: (a) receiving warfarin or any strong inhibitor of the cytochrome P450 family of enzymes responsible for metabolising most prescribed drugs; b) aggressive forms of CLL with central nervous system involvement; c) previous allogeneic or autologous stem-cell transplant. The ERG considers that these inclusion/exclusion criteria are appropriate.

A flow-chart of the participants in the MURANO trial was presented in CS pg 41. Of the 389 randomised patients in the trial, 382 (98%) received at least one dose of the assigned treatment, including 194 in the VEN+R arm and 188 in the BR arm. Twenty-eight patients withdrew from the trial: eight in the VEN+R group and 20 who were randomised to the BR group. The difference in withdrawal rates was significant (4% vs 10%, $p < 0.02$). However, the ERG would be more concerned if withdrawal rates were much higher in the VEN+R arm compared to BR, especially considering that BR is administered for a total of six 28-day cycles and VEN+R is given for two years.

After a median follow-up duration of 24.8 months, 78 of the 194 patients who received at least one dose of either venetoclax or rituximab remained on treatment, however, 68 participants already completed the two-year venetoclax treatment. Forty-eight patients in the VEN+R arm discontinued venetoclax with or without rituximab, including 10 patients who stopped following disease progression or relapse and 24 patients who discontinued treatment as a result of adverse events (AEs) (clarification response A7). Patients in the BR arm were also assessed and followed similarly as patients in the VEN+R arm. After a median follow-up duration of 22.1 months in the BR group, 154 of 188 patients who received at least one dose of either bendamustine or rituximab completed the treatment schedule. Expectedly, there were fewer discontinuations in the BR arm (n = 27) given the relatively shorter course of treatment. However, the main reasons for BR discontinuations were also disease progression or relapse (n = 6) and AEs (n = 11).

The baseline characteristics of patients enrolled in MURANO are reported in Table 3. Although it would appear that patients were seemingly healthy (as determined by CLL staging and ECOG

scores) entering into the trial, the ERG notes that there were no meaningful differences in demographic or disease characteristics between VEN+R or BR groups at baseline. The ERG requested clarification for Rai staging at diagnosis for 64 patients in the VEN+R group and 55 patients in the BR group who had not been accounted for. The company responded by providing Binet staging for these missing patients instead (clarification response A2). Although the degree of concordance between these staging systems remains uncertain, the ERG notes that the distribution of patients across the Binet stages are roughly comparable to patient distribution across the Rai stages, and are similar between VEN+R and BR groups. The ERG also requested a breakdown of the patients by country and geographical region in order to determine how applicable the findings were to the UK population. Although there were only 10 patients from the UK (six in VEN+R and four in BR), about two-thirds of the trial population were of European descent (130 in VEN+R and 131 in BR), which eased the ERG's concerns (clarification response A5). The ERG clinical expert also considers that the population of the MURANO trial was generalisable to UK population.

Table 3: Summary of baseline characteristics of MURANO patients

Period of enrolment	March 2014 to Sept 2015	
Characteristic	**VEN+R (n=194)**	**BR (n=195)**
Male n (%)	136 (70.1)	151 (77.4)
Age Median (min–max)	64.5 (28–83)	66.0 (22–85)
ECOG score of 0 / 1	111 (57.2) / 82 (42.3)	108 (55.7) /84 (43.3)
ECOG score of 1		
Rai staging Stage 0–II / Stage III–IV	88 (67.7) / 30 (23.1)	103 (73.6) / 18 (12.9)
Del(17p) status present	46 (26.6)	46 (27.2)
TP53 mutation status, n (%)		
N	192	184
Mutated	48 (25.0)	51 (27.7)
Unmutated	144 (75.0)	133 (72.3)
Del(17p) vs. TP53 mutation status, n/N (%)	171	158
Only del(17p)	24 (14.0)	18 (11.4)

TP53 mutation only	19 (11.1)	23 (14.6)
Del(17p) and TP53 mutated	22 (12.9)	22 (13.9)
Immunoglobulin heavy-chain variable (IGHV) Mutated	53 (29.4)	51 (28.3)
Risk status with regards to responsiveness to prior therapy, n (%)		
High	109 (56.2)	118 (60.5)
Low	84 (43.3)	75 (38.5)
Number of prior CLL therapy, n (%)		
1 previous line	111 (57.2)	117 (60.0)
2 previous lines	57 (29.4)	43 (22.1)
3 previous lines	22 (11.3)	34 (17.4)
>3 previous lines	4 (2.1)	1 (0.5)
Type of prior CLL therapies, n (%)		
Alkylating agent	182 (93.3)	185 (95.4)
Purine analogue	157 (80.5)	158 (81.4)
Anti-CD20 antibody	153 (78.5)	148 (76.3)
B-cell receptor inhibitors	3 (1.5)	5 (2.6)

4.3 Description and critique of company's outcome selection

The NICE scope lists the specified outcomes as:

- progression-free survival (PFS)
- overall survival (OS)
- response rates
- minimal residual disease (MRD) negative rate assessed in blood and bone marrow
- adverse effects of treatment
- health-related quality of life (HRQoL).

In the MURANO RCT, PFS was assessed by investigators (investigator-assessed PFS), which was the primary endpoint, and by an independent review committee (IRC-assessed PFS) and this was a secondary endpoint. In both cases, PFS was defined as the time from randomisation to the first occurrence of progression or relapse using the International Workshop on Chronic Lymphocytic Leukaemia (iwCLL) guidelines[16, 17] or death from any cause, whichever occurs first.

On Table 9 of the CS, the company has reported the protocol criteria for response based on 2008 iwCLL guidelines. These guidelines include parameters related to tumour load (lymphadenopathy, hepatomegaly, blood lymphocytes count, marrow infiltration) and to function of hematopoietic system or marrow (platelets and neutrophils counts, haemoglobin level).

On page 38 of the CS, the company has acknowledged that PFS can be affected by timing of assessments and can be prone to investigator bias but has stated that the use of strict criteria for response evaluation was implemented in the MURANO RCT. To evaluate disease status, patients were evaluated through computerised tomography (CT) scans of target lesions, blood counts and physical examinations of indicator lesions in up to six of the largest dominant nodes or tumour masses as well as in six extra-nodal lesions. The same was done for non-target lesions.

While the ERG agree that there was a strict protocol in place to assess disease status by investigators and that investigator-assessed PFS is a more relevant to the clinical practice, the ERG believe that IRC-assessed PFS was more preferable to investigator-assessed PFS as the former suggests that the outcome assessors were blinded to treatment allocation, reducing the potential for bias.

OS was defined as the time from randomisation to death from any cause.

To monitor HRQoL, the company used the EuroQoL five-dimension 3-level version (EQ-5D-3L) which was collected at regular intervals before progression, once at progression, and once at the first assessment following progression.

MRD negative rate was assessed through the clearance rate of MRD from blood or marrow samples. However, not all patients had both blood and bone marrow testing. Although the company reports a high level of concordance among patients who had both blood and bone marrow testing, the ERG is concerned that more patients had MRD peripheral blood testing than bone marrow because bone marrow is considered more sensitive than peripheral blood for MRD detection in CLL.[18]

Adverse events were graded using the National Cancer Institute Common Terminology Criteria for Adverse Events v4.0.

Overall, the outcomes selected in the CS are consistent with those identified by NICE as relevant to the decision problem.

4.4 Summary and Critique of MURANO Trial Statistics

The company's approach to trial statistics is presented in CS section B.2.4. Generally, statistical analyses entailed the use of stratified log-rank tests or stratified Cochran-Mantel-Haenszel tests, both of which were suitable for the design of the trial. Hazard ratios (HR) were obtained using stratified Cox proportional hazards (PH) models, however, no assessment of the proportional hazards assumption was made within the clinical effectiveness section of the company's submission.

The ERG reproduced a similar sample size calculation to that presented by the company and are satisfied that the trial was suitably powered to detect the specified difference in the primary outcome (HR of 0.66 in PFS). The results presented by the company were based on interim analyses planned after 140 events (75%) had occurred:[11] there were 146 reported events in the MURANO trial. The interim analyses were reviewed by an independent data monitoring committee, who recommended that the primary analysis be performed at this data cut-off. The final analysis was originally planned for the trial after 186 events had occurred. The interim analysis was also originally planned to be implemented 12 months after the final patient was enrolled into the study, however this was amended in version six of the study protocol.

For the primary outcome (investigator-assessed PFS), the company mention adjusting their significance level at 0.05 when performing a stratified log-rank test, however, no further detail was provided on this adjustment in their submission. Upon examining the clinical study report (CSR) provided by the company, the ERG discovered that the significance level at the primary endpoint was actually 0.0498, whereas a significance threshold of 0.002 was set for the interim analysis performed after 140 events had occurred. Nonetheless, as the interim analysis has become the primary analysis, the ERG do not believe this has any major consequence on the type-1 error rate of the trial outcomes.

The log-rank tests were stratified by del(17p) status, CLL risk status, and geographic region. The company also implemented a fixed sequence testing procedure which was not referred to in their submission. The following secondary endpoints were tested in the order presented:

- Complete response rate (CR) based on IRC assessment in all randomised patients (0.05 threshold, 2-sided)
- Overall response rate (ORR) based on IRC assessment in all randomized patients (0.05 threshold, 2-sided)
- OS in all randomized patients (0.0001 threshold, 2-sided)

Formal hypothesis testing would stop when one of the outcomes was not significant. The ERG is unsure why other secondary outcomes were not included in the fixed-sequence procedure, notably the proportion of patients achieving MRD-negativity. The final hypothesis test on OS is planned to be conducted 3 years after the final patient has been enrolled, and will use a 2-sided threshold of 0.0499, however this endpoint has not yet been reached. IRC-assessed PFS of patients with 17p deletion was originally included in the fixed sequence testing procedure, however this secondary outcome was excluded in version 4 of the statistical analysis plan. This secondary outcome, as well as the other secondary outcomes were tested at the 0.05 significance threshold which could have increased the likelihood of a Type 1 error.

Treatment allocation was performed using a block stratified randomisation procedure, which was deemed suitable by the ERG.

The ERG examined the approaches to trial statistics of the RESONATE and Study 116 trials, due to their importance in the indirect treatment comparison.

RESONATE (ibrutinib): This trial was assessed by an ERG during the TA429 appraisal of ibrutinib. The statistical analyses in the submission were based on Cox-models and log-rank tests, similar to the RESONATE study. The ERG for TA429 had no major concerns over the trial statistics.

STUDY 116 (IDELA+R): This trial was assessed by Warwick ERG during the TA359 appraisal of idelalisib in combination with rituximab. The approach to trial statistics was also similar to the RESONATE study, and entailed Cox-PH models and log-rank tests. The ERG of TA359 did not report any major concerns with the approach to trial statistics in Study 116.

Overall, the ERG has no major concerns over the approach to individual trial statistics of MURANO, RESONATE or STUDY 116.

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

4.5.1 Effectiveness

In this section, the ERG has summarised and critiqued the results from the MURANO trial.

The key results, including survival outcomes (PFS and OS) and response outcomes, are summarised in Table 4 and discussed in the following sections. In the table, the results are reported differently, some as number, some as %. There is little difference between investigators and IRC.

Table 4: Main survival outcomes

	VEN-R	BR
Number of patients	194	195
Median follow-up period	23.8 months	
Progression free survival (PFS): progression assessed by <u>investigators</u> (primary endpoint)		
Number of progressions or deaths	32	114
Median PFS (months)	Not reached	17
HR for progression or death (95% confidence interval (CI)) p-value	0.17 (0.11, 0.25) <0.0001	
Progression-free rates: % (95% CI)		
at 1 year	93 (NR)	73 (NR)
at 2 years	84.9 (79.1, 90.6)	36.3 (28.5, 44.0)
Progression free survival (PFS): progression assessed by <u>IRC</u> (secondary endpoint)		
Number of progressions or deaths	NR	NR
Median PFS (months)	Not reached	18.1
HR for progression or death (95% CI) p-value	0.19 (0.13, 0.28) <0.0001	
Progression-free rates: % (95% CI)		
at 1 year	NR	NR
at 2 years	82.8 (76.6-88.9)	37.4 (29.4-45.4)
Overall survival (OS)		
Number of deaths	NR	NR
Median PFS (months)	Not reached	Not reached
HR for death (95% CI) P value	0.48 (0.25, 0.90) NR	
OS rates: % (95% CI)		
at 1 year	NR	NR
at 2 years	91.9 (NR)	86.6 (NR)

NR: not reported in company submission; VEN-R: venetoclax rituximab; BR: bendamustine rituximab; IRC: independent review committee

4.5.1.1 Progression-free survival

Following a median of 23.8 months of controlled follow-up, the risk of progression or death was significantly lower in the VEN+R group compared to the BR group, irrespective of whether PFS was assessed by investigators (primary endpoint) (HR, 0.17; 95% confidence interval (CI): 0.11 to 0.25; $p < 0.0001$) or by an IRC (secondary endpoint) (HR 0.19; 95% CI: 0.13 to 0.28; $p < 0.0001$). These results were robust to sensitivity analyses conducted by the company.

4.5.1.2 Overall survival

The risk of death was significantly decreased in the VEN+R group compared to the BR group despite the limited duration of follow-up (HR, 0.48; 95% CI: 0.25 to 0.90; p = NA). However, OS results were still immature given that median OS was not reached in both arms.

4.5.1.3 Response outcomes including MRD outcomes

The rate of complete response (CR) or CR with incomplete hematologic recovery (CRi) was 18.6% higher ($p < 0.0001$) in the VEN+R arm compared to BR when assessed by investigators (see Table 5). However, there was no statistically significant difference in CR/CRi rates between VEN+R and BR when assessed by the IRC. On page 50 of the CS, the company has provided a reason for this discrepancy between the investigators and IRC indicating that there was a difference in the interpretation of residual adenopathy on CT especially regarding lesions measuring ≤30mm. The ERG clinical advisor agrees with the company's rationale.

Table 5: Main response outcomes including MRD outcomes

	VEN-R	BR
Response outcomes:		
Assessed by IRC		
CR / CRi: % (95% CI)	8.2 (NR)	3.6 (NR)
Difference on CR / CRi: % (95% CI) ; p-value	4.7 (-0.3 to 9.6); <0.081	
ORR: % (95% CI)	92.3 (87.6 to 95.6)	72.3 (65.5 to 78.5)
Difference on ORR: % (95% CI); p-value	20.0 (12.4 to 27.6); <0.0001	
Assessed by investigators		
CR / CRi	26.8 (NR)	8.2 (NR)
Difference on CR / CRi: % (95% CI) ; p value	18.6 (NR); <0.0001	
ORR: % (95%CI)	93.3 (88.8 to 96.4)	67.7 (60.6 to 74.2)
Difference on ORR: % (95% CI); p-value	25.6 (17.9 to 33.3); <0.0001	
Clearance rates of MRD:		
Based on peripheral blood samples		
At 9-months time point: n (%)	121 (62.4)	26 (13.3)
Absolute difference: % (95% CI) ; p-value	49.0 (40.4 to 57.6); NR	
At any time during the trial: n (%)	162 (83.5)	45 (23.1)
Absolute difference: % (95% CI); p-value	60.4 (52.3 to 68.6); NR	
Based on bone marrow aspirate		
At any time during the trial: n (%)	53 (27.3)	3 (1.5)
Absolute difference: % (95% CI); p-value	25.8 (19.0 to 32.6); <0.0001	

Overall response rate was improved (although non-significantly so) in the VEN+R group compared to the BR group, irrespective of whether ORR was assessed by investigators (absolute difference of 25.6%, 95% CI 17.9 to 33.3) or by an IRC (absolute difference of 20.0%, 95% CI 12.4 to 27.6).

Patients in the VEN+R group achieved higher clearance rates of MRD based on peripheral blood samples (absolute difference of 60.4%, 95% CI 52.3 to 68.6 at any time of the trial) and on bone marrow aspirate (absolute difference of 25.8%, 95% CI 19.0 to 32.6 at any time of the trial).

Although MRD assessments of bone marrow aspirates were only available for 29.6% (n = 115) of patients and peripheral blood MRD assessments available for 94.1% (n = 366), the company asserts that the level of concordance between MRD status in peripheral blood and bone marrow was 84.3% based on 108 pairs of post baseline samples across both treatment groups (82.5% for the VEN+R treatment group matching 85.3% for the BR treatment group). The ERG agrees with this assertion.

4.5.1.4 Health-related quality of life (HRQoL)

In the MURANO trial, HRQoL was measured using the EQ-5D-3L version questionnaire. On page 56 of the CS, it is indicated that only 35% of patients in the VEN+R group completed baseline patient-reported outcomes due to an undetected protocol error. Upon request, the company provides a breakdown of utility data by treatment arm, which revealed that patients in the VEN+R arm did not have a worse HRQoL than patients in the BR arm (Clarification Response A12). However, the ERG considers that this finding may have been influenced by the open-label nature of the MURANO trial. Overall, the ERG believes that the reliability of HRQoL outcomes is questionable.

4.5.1.5 Subgroup analyses

The company has presented a number of analyses by predefined subgroups in CS page 59 for the primary endpoint, investigator-assessed PFS.

These subgroups were:

- Age (<65 vs ≥65 yrs)
- CLL risk status (low vs high)
- Geographical region
- Number of previous therapies (1 vs 2 vs ≥3)
- Effect of most recent therapy
- Del(17p) status
- TP53 mutation status
- Baseline immunoglobulin heavy-chain variable (IGHV) mutation status

Results based on these pre-defined subgroups did not identify any subgroups more or less likely to benefit significantly from VEN+R. For instance, the risk of death or progression as assessed by the investigators was significantly higher in the VEN+R arm than the BR arm among R/R CLL patients with positive (HR 0.13, 95% CI 0.05 to 0.29) and negative (HR 0.19, 95% CI 0.12 to 0.32) 17p deletion status alike. Similarly, R/R CLL patients with TP53 mutation (HR 0.15, 95% CI 0.09 to 0.25) and non-mutation (HR 0.19, 95% CI 0.10 to 0.36) experienced significantly higher rates of death or progression in the VEN+R arm compared to the BR arm. Overall, the treatment benefit of VEN+R over BR was consistent across all subgroups.

4.5.2 Safety

Table 6 compares the safety of VEN+R and BR. Overall, there were more AEs in the VEN+R arm (n = 335) than in the BR arm (n = 255). Discontinuation rates due to AEs were also significantly higher in the VEN+R arm compared to BR (12.4% versus 5.9%, p = 0.03). However, it is not specified in the CS or CSR if AEs were treatment-related. The ERG also notes that the EMA is yet to ascertain the safety of VEN+R.

Grade 3 or 4 adverse events

Although the proportions of all patients with grade 3 or 4 AEs, defined using the National Cancer Institute (NCI) Common Terminology Criteria for Adverse Events (CTCAE) criteria (Protocol, pg 111), were significantly higher in the VEN+R arm compared to BR (82% versus 70.2%, P = 0.007), the only grade 3 or 4 AE with a significantly higher occurrence in VEN+R compared to BR was neutropenia (57.7% versus 38.8%, P = 0.0002). In this condition, the serum concentrations of white blood cells called neutrophils are decreased below the normal range, predisposing the patient to a number of infections. However, the ERG agrees that the low neutrophil count can easily be corrected if treated promptly; this is consistent with previous evidence analysing the safety of VEN+R.[19] The percentages of the other grade 3 or 4 AEs were either comparable between treatment arms (infections, anaemia, thrombocytopaenia, tumour lysis syndrome (TLS) and grade 3 or 4 AEs with less than 2% difference in incidence between VEN+R and BR) or significantly higher in the BR arm (febrile neutropaenia, infusion-related reaction and hypotension).

Serious adverse events (SAEs)

SAEs were broadly described as life-threatening or fatal according to the NCI CTCAE. Again, the proportions of SAEs were either similar between treatment arms or significantly higher in the BR arm. However, the ERG is unsure why the number of patients diagnosed with SAE pneumonia (n = 16) is greater than the number diagnosed with Grade 3 or 4 pneumonia in the VEN+R arm (n = 10). The ERG would expect fewer occurrences of SAEs compared to grade 3/4 AEs as is the pattern with other SAEs listed in CS Table 23.

Safety of VEN+R versus ibrutinib

The company has not compared the safety profile of VEN+R against any of the comparators in the scope. There are also no trials that directly compare AEs between VEN+R and ibrutinib or IDELA+R. However, the ERG clinical advisor suggests that the side effect profile of venetoclax is favourable compared to its key comparator ibrutinib. The ERG clinical advisor also suggests that the two-year stopping rule of VEN+R makes this intervention more attractive than ibrutinib which is administered indefinitely until disease progression.

Table 6: Summary of Adverse Events

Event	VEN+R (n=194)	BR (n = 188)	ERG-calculated p-values
Grade 3 or 4 AE — no. of patients (%)	159 (82.0)	132 (70.2)	0.01
Total no. of events	335	255	
Discontinuations due to AEs	24	11	0.03
Grade 3 or 4 AEs with at least 2% difference in incidence between groups — no. of patients (%)	130 (67.0)	104 (55.3)	0.02
Neutropenia	112 (57.7)	73 (38.8)	<0.001
Infections and infestations	34 (17.5)	41 (21.8)	0.29
Anaemia	21 (10.8)	26 (13.8)	0.37
Thrombocytopenia	11 (5.7)	19 (10.1)	0.11
Febrile neutropenia	7 (3.6)	18 (9.6)	0.02
Pneumonia	10 (5.2)	15 (8.0)	0.26
Infusion-related reaction	3 (1.5)	10 (5.3)	0.04
TLS	6 (3.1)	2 (1.1)	0.17
Hypotension	0	5 (2.7)	0.02
Hyperglycaemia	4 (2.1)	0	0.05
Hypogammaglobulinemia	4 (2.1)	0	0.05
SAEs — no. of patients (%)	90 (46.4)	81 (43.1)	0.52
SAEs with at least 2% incidence in either group — no. of patients (%)	47 (24.2)	76 (40.4)	<0.001
Pneumonia	16 (8.2)	15 (8.0)	0.92
Febrile neutropenia	7 (3.6)	16 (8.5)	0.04
Pyrexia	5 (2.6)	13 (6.9)	0.04
Anaemia	3 (1.5)	5 (2.7)	0.45
Infusion-related reaction	1 (0.5)	6 (3.2)	0.05
Sepsis	1 (0.5)	4 (2.1)	0.17
TLS	4 (2.1)	1 (0.5)	0.19
Hypotension	0	5 (2.7)	0.02
Fatal AEs	10 (5.2)	11 (5.9)	0.76

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

To reiterate, RESONATE and Study 116 were included as comparator trials in the MAIC.

Table 7 compares the study methods between MURANO and the comparator trials. The baseline characteristics of patients in MURANO, RESONATE and Study 116 are compared in Tables 100 (Appendix D1.1.8.2, pg 67) and 103 (Appendix D.1.1.8.3, pg 74) of the company's appendices, and are discussed in section 4.7.

To summarise, the RESONATE trial was a multicentre, open-label, phase 3 study in which 391 patients with R/R CLL or small lymphocytic lymphoma (SLL) were randomly assigned to receive daily oral ibrutinib until disease progression or toxicity occurs, whichever comes first, or weekly (and subsequently monthly) intravenous ofatumumab for up to 24 weeks. At baseline, a significantly higher proportion of patients in the ibrutinib group had bulky disease ≥ 5cm compared to the ofatumumab group (64% versus 52%, p = 0.04), and the median time from the last treatment received prior to enrolment in the trial was four months shorter in the ibrutinib arm compared to the ofatumumab arm (8 mo versus 12 mo, p = 0.02). However, there were no other significant differences between the two groups at baseline. The primary endpoint was duration of PFS as assessed by an IRC, whereas OS duration and ORR were key secondary endpoints. The results show ibrutinib to be superior to ofatumumab. At a median follow-up of 9.4 months, the median PFS duration had not been reached in the ibrutinib arm, as compared to 8.1 months in the ofatumumab arm (HR 0.22, p < 0.001). Similarly, ibrutinib significantly improved OS (HR 0.43, p = 0.005) and ORR (42.6% versus 4.1%, p < 0.001). The statistical analyses in the trial were based on Cox- proportional hazard models and log-rank tests, which the ERG deems appropriate. The ERG also agrees with the company's quality assessment of RESONATE as presented in Table 110 of CS Appendix D1.3, and judges the trial to be of good quality.

Study 116 was a randomised, double-blind, placebo-controlled, phase 3 trial in which 220 patients with decreased kidney and bone marrow function were randomised to receive rituximab in combination with either idelalisib (IDELA+R) or placebo (placebo + rituximab). Although the baseline characteristics, as presented in the published trial, were comparable between treatment arms, the ERG is unsure how similar at baseline the proportions of patients with kidney and bone marrow diseases (or any other co-existing conditions) are between idelalisib and placebo. The primary endpoint was PFS and the secondary endpoints included OS and ORR. An independent data and safety monitoring board stopped the trial at the first pre-specified interim analysis following results of the overwhelming efficacy of idelalisib: median PFS was 5.5 months in the

placebo group but had not been reached in the idelalisib group (HR 0.15, $p < 0.001$). The statistical analyses in the trial were based on Cox- proportional hazard models and log-rank tests, which the ERG deems appropriate. The ERG also agrees with the company's quality assessment of RESONATE: although Study 116 is a double-blind randomised trial, the risk of selection bias in this study may be high as details of the randomisation procedure and allocation concealment are not reported (CS Appendix D1.3, Table 110).

Table 7: Comparison of study methods across the MAIC trials

	MURANO	RESONATE	STUDY 116
Comparators and dose	Venetoclax (ramped up to 400 mg per day, oral) + rituximab vs Bendamustine + rituximab	Ibrutinib (420 mg once daily, oral) vs Ofatumumab	Idelalisib (150 mg twice daily, oral) + rituximab vs Placebo + rituximab
Location	109 sites in 20 countries including US, Canada, Australia, New Zealand, and countries in Europe and Asia	67 sites in the United States, Australia, and seven European countries	90 Centres in US and Europe
Trial Design	1:1 multicentre randomised, open-label, phase 3 trial	1:1 multicentre randomised, open-label, phase 3 trial	1:1 multicentre randomised, double blind phase 3 trial
Eligibility Criteria	18 years of age or older Diagnosed with R/R CLL that also required therapy Received one to three previous treatments (including one or more chemotherapy-containing regimens) ECOG score of 0 or 1 Adequate bone marrow, kidney, and liver function Patients who had received previous treatment with bendamustine were eligible provided that the duration of response after the treatment was at least 24 months.	Patients with previously treated CLL or SLL who require therapy were eligible Unsuitable for purine analogue therapy (e.g. patients with short progression-free interval after chemoimmunotherapy, co-existing illnesses, 70 years of age or more, or presence of 17p deletion). ECOG score of 0 or 1 Absolute neutrophil count of at least 750 cells per microliter Platelet count of at least 30,000 cells per microliter	Patients with CLL that had progressed within 24 months after their last treatment Unsuitable for cytotoxic therapy (e.g. severe neutropenia or thrombocytopenia caused by cumulative myelotoxicity from previous therapies, an estimated creatinine clearance of less than 60 ml per minute, or a CIRS score on the Cumulative Illness Rating Scale (CIRS) of 6 or more for coexisting illnesses not related to CLL

		Adequate liver and kidney function Patients requiring warfarin or strong CYP3A4/5 inhibitors were excluded.	Previous treatment must have included either a CD20 antibody-based regimen or at least two previous cytotoxic regimens.
Outcomes of interest	PFS OS IRC PFS ORR CR MRD clearance Event-free survival (EFS) Duration of Response Time to next treatment	PFS OS ORR	PFS OS ORR CR Lymph Node Response HRQoL
Crossover details	Crossover was not permitted in the trial design.	Patients on Ofatumumab were able to switch to ibrutinib following disease progression	Patients on placebo were able to switch to idelalisib following disease progression.
Randomisation strata	Presence or absence of chromosome 17p deletion Responsiveness to previous therapy Geographic region	Resistance to purine analogue therapy (defined as no response or a relapse within 12 months after the last dose of a purine analogue) Presence or absence of chromosome 17p deletion	Presence or absence of 17p deletion and/or TP53 mutation Presence or absence of unmutated IGHV
Subgroups	Age (<65y vs ≥65y) CLL risk (low vs high)[a] Geographic Region (North America vs Asia vs Western Europe vs Central/Eastern Europe vs Australasia) Number of previous therapies (1 vs 2 vs ≥3) Presence or absence of chromosome 17p deletion TP53 mutation status	Age (<65y vs ≥65y) Gender (male vs female) Race (white, non-white) Geography (Europe vs USA) Rai Stage (0-2 vs 3-4) ECOG Score (0 vs 1) Bulky disease (<5cm vs ≥5cm)	IGHV (mutated vs unmutated) Presence or absence of 17p deletion and/or TP53 mutation Presence or absence of 17p deletion Gender (male vs female) Age (≤65y vs >65y)

	IGHV mutation status Effect of most recent therapy (relapse vs refractory)	Number of previous treatments (<3 vs ≥3) Presence or absence of chromosome 17p deletion Presence or absence of 11q22.3 deletion Baseline β_2 microglobulin level (≤ 3.5mg/L vs >3.5mg/L Resistance to purine analogue therapy (yes *vs* no)	

[a] High-risk CLL status was defined as any of the following: presence of 17p deletion, no response to front-line chemotherapy-containing regimen, relapsed disease with 12 months of chemotherapy alone, or relapsed disease within 24 months of chemoimmunotherapy.
CIRS, Cumulative Illness Rating Scale: The CIRS score ranges from 0 to 56, with higher scores indicating an increased number or greater severity of coexisting illnesses.

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

Using individual patient data (IPD) from the RESONATE and HELIOS trials, Hillmen and colleagues published an indirect comparison of ibrutinib-BR combination versus BR versus single agent ibrutinib.[20] However, the company states in CS section B.2.9.4 that IPD were neither available for RESONATE nor for Study 116. Hence, the MAIC entailed comparison of aggregate data from RESONATE and Study 116 with IPD from MURANO.

Clinical trial selection

The RESONATE and MURANO trials were both open-label with similar inclusion/exclusion criteria, whereas Study 116 was a double-blind trial with contrasting criteria: while patients with adequate kidney and bone marrow function were eligible for inclusion in MURANO and RESONATE, such patients were excluded from Study 116. Table 8 and Table 9 compare other characteristics between the trials. As shown, there were cross-trial differences in a number of baseline characteristics including age, Rai stage, ECOG score, bulky disease status and Beta-2 Microglobulin concentration. Without IPD from the comparator trials in the MAIC, the ERG is concerned that there may still be residual unobserved differences and potential sources of bias even after matching.[15] Nonetheless, the ERG regards the implementation of the MAIC as reliable.

Identification of outcome measures

The primary end-points in MURANO and RESONATE were assessed differently: investigator-assessed PFS was the primary outcome in the MURANO trial, whereas PFS was assessed by the IRC in the RESONATE trial. In the indirect comparison of these trials, the ERG considers that the IRC-assessed PFS IPD in MURANO (a secondary end-point in the trial) should have been reanalysed to match the IRC-assessed PFS in RESONATE.[15] However, as illustrated in CS Table 20 (pg 70), the outcome measure used was investigator-assessed PFS. The MURANO IPD matched the primary outcome measure used in Study 116 (CS Table 20, pg 70).

Matching trial populations

Fifty-six patients in the MURANO trial (25 in the VEN+R arm versus 31 in the BR arm) who had an ECOG score of > 1 or received prior B-cell receptor inhibitor therapy were excluded from the indirect comparison because these patients would have been ineligible for the published RESONATE trial. Similarly, 54 patients in the MURANO trial (24 in VEN+R versus 30 in BR) who would not have been eligible to be included in Study 116 were excluded from the indirect comparison.

To adjust for residual cross-trial differences, patients in the MURANO trial were weighted such that their weighted mean baseline characteristics matched those reported for the RESONATE and Study 116 trials. While previous evidence supports this approach to matching,[15] the ERG is concerned about the marked deviation of the matched sample characteristics (such as age, Rai stage, bulky disease status, prior therapy status, ECOG score, and Beta-2 microglobulin concentration) and sample size from the original MURANO trial population (N = 194 in VEN+R arm). It is also unclear what informs the arbitrary significance threshold of 0.25 used for selecting variables/effect-modifiers on which trials were matched. However, the ERG acknowledges that the trials were matched on the relevant prognostic factors of R/R CLL.

Network of evidence

A schematic of the evidence network for the relevant comparators in the MAIC is presented in CS Figure 13. The evidence network shows there was no common comparator connecting all the

treatments in the included trials (VEN+R, ibrutinib, IDELA+R). Hence, the evidence network was disconnected and unanchored MAIC analyses were performed to estimate the relative effectiveness of VEN+R over ibrutinib (and IDELA+R) and inform the base-case HRs for PFS and OS (clarification response A10).

However, the company acknowledges that there is a higher risk of residual bias associated with performing an unanchored MAIC, and sought to perform an exploratory anchored MAIC analysis for testing the robustness of the unanchored MAIC results. An anchored MAIC is a standard indirect treatment comparison with a common comparator for the treatments in the network, and the company uses evidence from the MURANO (VEN+R versus BR) and HELIOS (ibrutinib+BR versus placebo+BR)[21] trials to perform this exploratory analysis because BR is the common comparator in both trials. The ERG agrees with this approach to sensitivity analysis, but disagrees that the effect estimates from the unanchored MAIC were consistent with the anchored MAIC results. More so, ibrutinib monotherapy, and not ibrutinib+BR, is the specified comparator in the final scope and decision problem. The company justifies using ibrutinib+BR in the anchored MAIC by citing Hillmen and colleagues who found single agent ibrutinib to be as effective as ibrutinib+BR for treating patients with R/R CLL.[20]

Table 8: Baseline characteristics of the trial populations in the MURANO and RESONATE trials before and after matching

Characteristics	Before matching		After matching	
	VEN+R MURANO	Ibrutinib RESONATE	VEN+R MURANO	Ibrutinib RESONATE
	(N=169)[a]	(N=195)	(N=62)[b]	(N=195)
Age ≥65	50.89%	60.51%	60.51%	60.51%
Rai stage III-IV	27.22%	55.90%	55.90%	55.90%
Bulky disease ≥5cm	43.79%	63.59%	63.59%	63.59%
Prior therapy >1	43.79%	82.05%	82.05%	82.05%
Chromosome 11q del	35.50%	33.16%	33.16%	33.16%
Chromosome 17p del	27.22%	32.31%	32.31%	32.31%
ECOG=1	45.56%	59.49%	59.49%	59.49%

IGVH=Mutated	29.59%	26.87%	26.87%	26.87%
β2-microglobulin>3.5 mg/L	64.50%	83.71%	83.71%	83.71%
Prior Purine Analog	80.47%	85.13%	85.13%	85.13%
Prior AntiCD20	73.96%	93.85%	93.85%	93.85%

[a] 25 patients with prior BCRi therapy, ECOG>1, and no central lab measurement for assessing del(17p) status were excluded from the VEN+R IPD population (N = 194) before matching. [b] About two-thirds of the VEN+R IPD population were unmatched to the ibrutinib arm of RESONATE. The ERG deemed the comparator arms in the trials (BR and ofatumumab) irrelevant to the table.

Table 9: Baseline characteristics of the trial populations in the MURANO and Study 116 trials before and after matching

Characteristics	Before matching		After matching	
	VEN+R MURANO	IDELA+R Study 116	VEN+R MURANO	IDELA+R Study 116
	(N=170)	(N=110)	(N=53)	(N=110)
Age ≥65	50.59%	80.91%	80.91%	80.91%
Rai stage III-IV	27.06%	67.37%	67.37%	67.37%
Prior therapy >1	56.47%	75.00%	75.00%	75.00%
Chromosome 11q del	35.88%	34.00%	34.00%	34.00%
Chromosome 17p del	27.06%	23.64%	23.64%	23.64%
IGVH=Mutated	29.41%	17.27%	17.27%	17.27%
β2-microglobulin>3.5 mg/L	64.12%	85.45%	85.45%	85.45%

[a] 24 patients with prior BCRi, ECOG>1, and no central lab measurement for assessing del(17p) status were excluded from the VEN+R IPD before matching. [b] About two-thirds of the VEN+R IPD population were unmatched to the IDELA-R arm of Study 116. The ERG deemed the comparator arms in the trials (BR and ofatumumab) irrelevant to the table.

Results from the MAIC analyses

The results from MAIC comparisons undertaken by the company are presented in CS Tables 20 and 21. The ERG has reviewed those regarding the VEN+R vs ibrutinib comparison given that single-agent ibrutinib has been acknowledged as the most relevant comparator to VEN+R: the ERG clinical advisor confirms that ibrutinib is considerably more effective than IDELA+R and is better tolerated. Based on adjusted comparisons, VEN+R is thought to reduce the risk of

progression or death compared to ibrutinib, although the difference is not statistically significant (PFS HR [redacted] 95% CI [redacted]); regarding the OS outcome, VEN+R was found to reduce the risk of death compared to ibrutinib, this reduction reached statistical significance (OS HR [redacted] 95% CI [redacted]).

The ERG was surprised by the magnitude of this result suggesting a [redacted]% reduction for the risk of death with VEN+R relative to ibrutinib. The magnitude of this benefit is in marked contrast to the CS MAIC results for PFS where VEN+R reduces the risk of progression or death by only [redacted]% compared to ibrutinib (this difference is not statistically significant).

For most RCTs conducted on cancer drugs, except those comparing immune checkpoint inhibitors to conventional chemotherapy treatment, there is usually a notable correlation between PFS and OS indicating that a positive benefit in PFS should translate into a positive benefit in OS, in other words, PFS is often thought to be a valid surrogate outcome to OS.[14] This is one of the reasons why in a number of cancer trials undertaken in people with early stage/moderately advanced disease stage, PFS is usually taken as primary endpoint, while OS, is chosen as secondary endpoint. Based on recent RCTs for drugs tested in patients with R/R CLL, one can observe the correlated trend between PFS and OS benefits (Table 10): a large benefit on PFS (low HR) seems to translate into a lower benefit (higher HR) in OS.

Table 10: Comparison of PFS and OS outcomes in R/R CLL

Study	**Treatment 1**	**Treatment 2**	**PFS HR** $_{1 \text{ vs } 2}$	**OS HR** $_{1 \text{ vs } 2}$
HELIOS[21]	Ibrutinib+BR	BR	0.20	0.63
MURANO[11]	VEN+R	BR	0.19	0.48
RESONATE[12]	Ibrutinib	Ofatumumab	0.22	0.43
Company's MAIC	VEN+R	Ibrutinib	[redacted]	[redacted]

However, these observed relationships between PFS and OS are at odds with the results of the company's MAIC, where a moderate (non-significant) reduction of the risk of progression or death translates into a very high reduction for the risk of death. A similar relationship has

previously not been observed and the ERG believes that nothing in the mechanism of action of VEN+R could explain these apparently incoherent and illogical results.

A crude indirect comparison between ibrutinib and BR using the MAIC results and VEN+R as a common comparator would suggest that the risk of death is reduced by approximately [redacted] with BR compared to ibrutinib (the OS HR for VEN+R vs BR is [redacted] while the MAIC calculated OS HR for VEN-R vs IBRU is [redacted]), which again appears to be implausible and contrasts with ibrutinib becoming the gold standard for treating people with R/R CLL since its recommendation in 2016.[9]

In the cost-effectiveness section, the ERG will further demonstrate the non-plausibility of OS HRs estimates from the MAIC by examining the predicted life expectancy for ibrutinib obtained through the cost-effectiveness model that used results from the MAIC. Given the OS HR from the MAIC which were deemed implausible, the ERG requested at clarification stage the set of data used by the company to undertake the analyses. The ERG used the data and MAIC code provided to reproduce and critique the MAIC performed by the company.

Critique of the MAIC Implementation

To reiterate, a MAIC can be used to compare two treatments when IPD is available for one treatment of interest, and summary data available for another treatment of interest. Either through the use of a common comparator (anchored) or not (unanchored), the MAIC estimates the efficacy of the treatment with IPD available in the population of the treatment with summary data. This is a cause for concern, as the company have estimated that the relative efficacy of VEN+R compared to ibrutinib in the population of the RESONATE trial through estimation of a hazard ratio, and assumed that the relationship will be identical in the MURANO trial population. The company have not discussed this assumption and the potential flaws. As previously demonstrated by AbbVie and Novartis, a MAIC conducted on the same two treatments, but from different perspectives can yield different estimates of relative efficacy (e.g. depending on which treatment you have IPD for, and numbers after matching).[22, 23] Hence, it is important to carefully consider the population of interest, and may not be appropriate to assume generalisability of a relative treatment effect from one trial population to another. It is also evidence that it is unlikely that all prognostic and treatment-effect modifiers are completely accounted for.

In an unanchored MAIC, it is important to include both prognostic and treatment modifiers, in order to allow adjustment for differences in trial population.[24]

In the company's selection of covariates, they specify a threshold of 0.25 for p-values of tests of prognostic factors and of interaction with treatment effect in MURANO. The ERG acknowledges that there is precedence for applying a 0.25 significance threshold when selecting a complete set of potential predictors,[25] however there are concerns that this may lead to the inclusion of variables that are having an interactive effect only by chance, without any true interactive effect with treatment. The additional concern of the ERG is the dichotomisation of several continuous or categorical variables, resulting in a potential large loss of information. The ERG understand that this was likely done to reduce the number of categories matched, thus increasing the sample size; however, this could result in, for example, a participant aged 65 being assumed equal to a participant aged 85, yet they will likely have considerably different life expectancy. With the dichotomised variables containing heterogeneous populations, there is no guarantee that the distribution of these variables is well matched after performing the MAIC.

The ERG scrutinised the MAIC approach conducted by the company, to verify that there were no major mistakes which could explain the implausible HR. The ERG found one error in the data extraction from RESONATE relating to the β2 microglobulin > 3.5mg/litre proportion. The correct proportion is 153/195 and not 298/356[12]. When the unanchored adjusted MAIC is re-run, the impact is minor (sample size of the VEN+R IPD 62 to 61; [redacted]

The covariates that met the prognostic criteria for association can be found in Table 11, alongside covariates that met the treatment interaction threshold.

The company examined other trials which presented data on treatment interactions, also shown in Table 11.

The company then reached the conclusion that the variables which met either of the following criteria would be considered as effect modifiers in their MAIC:

- MURANO variables with association $p<0.25$ when interacted with treatment.
- Some evidence of potential effect modifying status in comparator trial publication.

The ERG are concerned that some relevant prognostic factors may not have been included in the economic model, as they are not specifically considered in these criteria. They would only have been included if they were also a treatment-effect modifier. The ERG are also concerned about the lack of inclusion of modifiers which appear to meet the company's inclusion criteria. These were absolute lymphocyte count (ALC), creatinine clearance (CRCL), response duration of recent therapy, refractory to last anti-leukaemia therapy, BCRi and ZAP70 expression. It is possible that this is down to a lack of corresponding data in the comparator trials, however this is not discussed by the company. The ERG is concerned that despite the matching, there may remain considerable imbalances between excluded variables reported in Table 11 and other unmeasured variables, potentially biasing the analysis and contributing to the implausible estimates of treatment effect. In addition, the criteria have been selected based on their influence on the PFS outcome, yet are used in both PFS and OS MAICs. Whilst the immaturity may have prevented an OS based analysis, the company do not seem to have considered this approach, and assumed a direct relationship between OS and PFS.

Table 11: Comparison of potential MAIC factors from company's search

	MURANO prognostic modifiers	MURANO treatment effect modifiers	External study treatment effect modifiers	Covariates included for matching by company
Variables	• **Age** • Hispanic ethnicity • TLS risk • **bulky disease** • risk status • **central lab measurements for del(17p)** • 12 trisomy • chromosome 13 deletion • CRCL • **TP53 mutation** • **IGHV mutation** • refractory to last chemo-containing therapy • refractory to last anti-leukaemia therapy, • fludarabine refractory • **number of prior CLL treatments** • **prior purine analogue agent** • **prior anti-CD20** • time from first diagnosis • time from last prior therapy to randomization • time to randomization	• **Age** • **ECOG**, • **bulky disease** • ALC • **chromosome 11q deletion** • CRCL • **Beta-2 microglobulin** • **IGHV mutation** • response duration to recent therapy, • refractory to last anti leukaemia therapy • **number of prior CLL therapies**, • **prior purine analogue agent** • prior BCRi.	• **Age**, • **Rai stage** • **ECOG** • **Chromosome 11q deletion** • **IGHV** • ZAP70 expression • **number of prior therapies** • **Beta-2 microglobulin** • **del(17p) or TP53 mutation**.	• Age • Rai Stage • Bulky Disease • Number of prior treatments • Chromosome 11q deletion • Del(17p) or TP53 mutation • ECOG • IGHV Mutation • Beta-2 microglobulin • Prior Purine • Prior Anti CD20

	from relapse since last line of treatment.			
Included in Matching	8/20	8/13	8/9	11/11

Bold indicates variable was included in company's matching.

4.8 Additional work on clinical effectiveness undertaken by the ERG

For the purpose of cost-effectiveness modelling, the ERG has proposed another method to estimate the relative benefit of VEN+R compared to ibrutinib given the implausible OS findings obtained from the MAIC.

The ERG agrees with the company's network of evidence for drugs used in R/R CLL presented in CS Figure 14, which suggests that there is no sufficient evidence to indirectly compare ibrutinib to VEN+R using results from RCTs.

However, the ERG has identified an abstract by Hillmen et al.[20] that compared single-agent ibrutinib to BR. This abstract was cited in the CS but the company did not use the results presented from this abstract for the purpose of comparing ibrutinib to BR. In this study, the authors use IPD data from the RESONATE and HELIOS RCTs to compare the efficacy of ibrutinib against BR after adjusting for a number of covariates, namely age, gender, Rai staging, ECOG score, del(11q) status, refractory status, number of prior lines of therapy, bulky disease, IGVH status. Results from this indirect comparison are reported in Table 12.

Table 12: Indirect comparison of ibrutinib versus BR

Study	**Treatment 1**	**Treatment 2**	**PFS HR** $_{1 \text{ vs } 2}$	**OS HR** $_{1 \text{ vs } 2}$
Hillmen et al. (2015)[20]	BR	Ibrutinib	7.52 (95% CI 4.72- 11.99)	2.24 (95% CI 1.14 -4.4)

Although the Hillmen et al. (2015)[20] results have not been obtained from a direct comparison, the use of IPD and appropriate methods of adjustment was deemed by the ERG to provide reasonable estimates of the ibrutinib vs BR comparison. Therefore, the ERG has decided to undertake

exploratory analyses to provide more robust estimates for the key clinical effectiveness outcome measures between ibrutinib and VEN+R. This was done using BR as common comparator.

The ERG compared hazard ratio (95% CI) estimates for PFS and OS across these two studies. For PFS outcomes, we used estimates obtained from IRC analyses. We used the package '*network*' in Stata 15[26] to conduct a network meta-analysis (NMA). Because this package operates in a frequentist paradigm, there was no need to perform sensitivity analysis on prior distributions. Given that the network was very sparse, we used a fixed-effects model. We used a common heterogeneity model, where the between-studies variance is assumed equal across comparisons. Since there was no mixed (direct + indirect) comparisons between interventions, there was no need to check networks for inconsistency. We did not present any rankograms or surface under the cumulative ranking curve (SUCRA) scores for these interventions.

PFS network meta-analyses

The data we used for the NMA for PFS are presented in Table 13.

Table 13: Data used in the ERG's NMA for PFS

Study	**Year**	**Treatment 1**	**Treatment 2**	**PFS_HR$_{1vs2}$**	**PFS_HR_ LCI$_{1vs2}$**	**PFS_HR_ UCI$_{1vs2}$**
Murano	2018	VEN+R	BR	0.19	0.13	0.28
Hillmen RESONATE+HELIOS	2015	BR Ibrutinib	Ibrutinib BR	7.52 0.13	4.72 0.083	11.99 0.211

LCI – lower confidence interval; UCI – upper confidence interval

The network of interventions is presented in Figure 1.

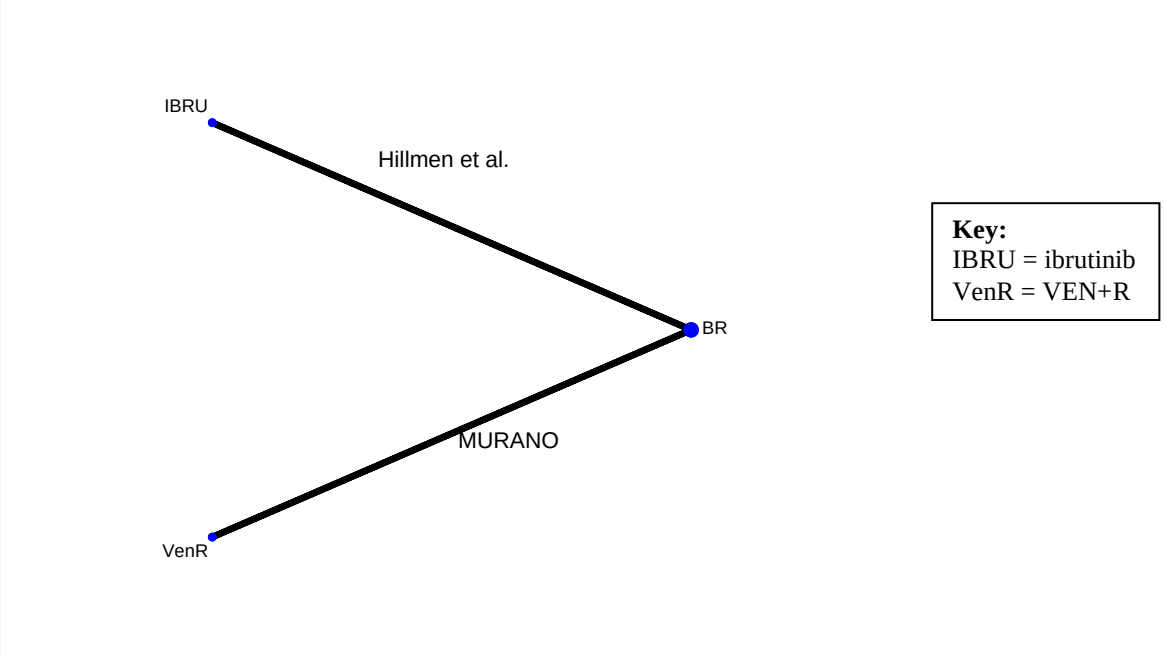


Figure 1: Network of interventions

Following the NMA, the HR for progression or death of VEN+R relative to ibrutinib is 1.43 (95% CI 0.78-2.61).

OS network meta-analyses

The data we used for the NMA for OS are presented in Table 14.

Table 14: Data used in the ERG's NMA for OS

Study	**Year**	**Treatment 1**	**Treatment 2**	**OS_HR$_{1vs2}$**	**OS_HR_ LCI$_{1vs2}$**	**OS_HR_ UCI$_{1vs2}$**
Murano	2018	VEN+R	BR	0.48	0.25	0.9
Hillmen	2015	BR	Ibrutinib	2.24	1.14	4.4
RESONATE+HELIOS		Ibrutinib	BR	0.45	0.23	0.88

LCI – lower confidence interval; UCI – upper confidence interval

The network of interventions for OS NMA is similar to that for PFS. Following the NMA, the HR for death of VEN+R relative to ibrutinib is 1.08 (95% CI 0.42-2.73).

Face-validity check and limitations

In Table 15, the ERG has summarised the PFS and OS estimates for the indirect comparison of VEN+R to ibrutinib using either the MAIC reported in the CS, or the ERG's exploratory NMA.

Table 15: Comparison of PFS and OS outcomes in R/R CLL using the MAIC or the ERG's exploratory NMA

Study	**Treatment 1**	**Treatment 2**	**PFS HR $_{1 \text{ vs } 2}$**	**OS HR $_{1 \text{ vs } 2}$**
Company's MAIC	VEN+R	Ibrutinib	[redacted]	[redacted]
ERG's NMA			1.43 (0.78-2.61)	1.08 (0.42-2.73)

There is a considerable difference between the company's and the ERG's estimates regarding the performance of VEN+R relative to ibrutinib. There is no formal argument to prefer the ERG's estimate for PFS rather than that of the company. However, the ERG believes that the estimates for both PFS and OS appear consistent with the idea that a benefit observed on PFS is associated with a lower benefit on OS. Moreover, when applied to the economic model, the ERG's estimates does not lead to implausible results with PFS exceeding OS for ibrutinib (see CS Figure 24).

We show in the cost-effectiveness section that using the ERG's NMA HR for OS in the model leads to an extrapolated life expectancy which is much more consistent with the predicted mean survival using reconstructed IPD.

The ERG acknowledges the exploratory nature of our analyses since we did not conduct a full systematic review to search for potential sources of additional of information. Furthermore, our NMA may seem simplistic because we cannot assess whether the transitivity assumption does hold.

4.9 Conclusions of the clinical effectiveness section

The ERG recognises the dearth of comparator studies relevant to the final scope and company's decision problem, and acknowledges the RESONATE and Study 116 trials as appropriate sources of aggregate data for comparison against IPD from the MURANO trial. The RESONATE trial investigated the efficacy of the more relevant comparator of VEN+R (single-agent ibrutinib) and matched better with the MURANO trial. The methods used in matching trial populations have been previously validated; however, the ERG is concerned about the imprecise estimates of the treatment effect of VEN+R (confidence intervals of HRs for PFS and OS were wide) as well as the implausible HRs for OS. Additional work undertaken by the ERG indirectly comparing estimates of the treatment effect of VEN+R from the MURANO trial against single-agent ibrutinib from Hillmen and colleagues[20] supports the ERG's position.

5 COST EFFECTIVENESS

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

5.1.1 Objectives and search strategy

The CS states on pg 82 that a systematic literature search was conducted to identify studies that assessed the cost-effectiveness of interventions for VEN+R and its appropriate comparators. The scope of the review was broadened to include all interventions in R/R CLL. Two other systematic reviews, aimed at identifying HRQoL data and relevant cost and resource use data for England and Wales that could be used in the company's economic model, are briefly described on pgs 119 and 130. The company provided an appropriate description of the cost-effectiveness, the HRQoL and the cost and health care resource use systematic reviews and details of the different search strategies were reported in Appendices G, H and I, respectively. In brief, the company searched MEDLINE, EMBASE, Econlit, Cochrane library including the NHS Economic Evaluation Database and HTA databases. Manual searches were also performed on seven conference proceedings websites and these searches were restricted to the last three years. In addition, reference lists of included papers were also consulted and for the HRQoL and cost and resource use reviews, previous NICE submissions in CLL were assessed. Original searches were carried out on 8 July 2017. Although these searches were updated on 30 April 2018, a limit to records with a publication date between 2014 and 2017 was applied. The search strategies were appropriate. The ERG has undertaken targeted searches to check for recent 2018 publications and has not identified any further cost-effectiveness studies, mainly due to the scarcity of evidence in this area.

5.1.2 Inclusion/exclusion criteria used in the study selection

The CS on pg 83-84 (CS table 26) tabulated the inclusion and exclusion criteria for the systematic reviews of economic evaluations which used the population, intervention, comparator, outcome (PICOS) framework and included: population, intervention/comparator, outcomes, study design type, publication type, and language. The selection criteria limited studies to those in adult patients 18 years or older, those with established R/R CLL including del(17p) R/R CLL patients, and studies published in English language. The study selection seemed appropriate. A similar inclusion/exclusion criteria was used for the HRQoL and cost and resource use reviews, however,

there were no restrictions applied on the type of interventions or type of comparators for these two reviews.

5.1.3 Included studies

CS Figures 18, 66 and 67 provided the flow diagrams for the cost-effectiveness, HRQoL, and cost and resource use systematic reviews, respectively. The cost-effectiveness search included 29 studies and 27 studies were excluded with complete references and reasons provided in Appendix G. Likewise, the HRQoL search included 13 studies and 20 studies were excluded with complete references and reasons provided in Appendix H; and the cost and resource use search included 16 studies and 14 studies were excluded with complete references and reasons provided in Appendix I.

The CS did not state whether the studies were independently assessed by two reviewers. Quality assessment for the cost-effectiveness studies was conducted by the company using the Drummond checklist[27] however, a more update checklist such as the CHEERS checklist[28] would have been more appropriate and it would have also been beneficial to have summary of the quality assessment.

To summarise, no cost-effectiveness studies assessing VEN+R for treating patients with relapsed or refractory CLL were identified.

5.1.4 Conclusions

The company did not provide a formal conclusion from the data available of the three systematic reviews: cost-effectiveness, HRQoL and cost and resource use.

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

5.2.1 NICE reference case checklist

Attribute	Reference case and TA Methods guidance	Does the de novo economic evaluation match the reference case
Comparator(s)	Therapies routinely used in the NHS. Including technologies regarded as current best practice for the two populations	Ibrutinib as an option for first line treatment of del(17p)/TP53 patients and second line treatment of non-del(17p)/TP53 patients. Idelalisib + rituximab for treatment of R/R CLL.
Patient group	As per NICE final scope	1. Patients with relapsed CLL - a CLL patient who previously achieved a CR or partial response/remission (PR), but after a period of six or more months demonstrates evidence of disease progression; 2. Patients with refractory CLL – a CLL patient who has progression within six months of the last anti-leukemic therapy R/R CLL population is split into two subgroups: a. patients with del(17p) and/or TP53 mutation

Attribute	Reference case and TA Methods guidance	Does the de novo economic evaluation match the reference case
		b. patients with non-del(17p) and/or TP53 mutation
Perspective costs	NHS & Personal Social Services	Yes
Perspective benefits	All health effects on individuals	Yes
Form of economic evaluation	Cost-effectiveness analysis	Cost-effectiveness analysis (Cost per quality-adjusted life year (QALY))
Time horizon	Sufficient to capture differences in costs and outcomes	Yes (lifetime duration – approximately 30 years)
Synthesis of evidence on outcomes	Systematic review	Data are drawn from one study: MURANO trial
Outcome measure	Quality-adjusted life years	Yes
Health states for QALY	Described using a standardised and validated instrument	Yes. Health states were evaluated using EQ-5D-3L data collected from MURANO trial
Benefit valuation	Time-trade off or standard gamble	The standard UK EQ-5D tariff is used, which is based upon time-trade off
Source of preference data for valuation of changes in HRQoL	Representative sample of the public	Yes
Discount rate	Annual rate of 3.5% on both costs and health effects	Yes
Equity	An additional QALY has the same weight regardless of the other characteristics of the	Yes

Attribute	Reference case and TA Methods guidance	Does the de novo economic evaluation match the reference case
	individuals receiving the health benefits	
Probabilistic modelling	Probabilistic modelling	Yes
Sensitivity analysis		A range of sensitivity and scenario analyses is presented

The cost-effectiveness evidence submitted by the company appears to satisfy the NICE reference case, and the decision problem defined in the scope.

5.2.2 Model structure

The company presented a *de novo* partitioned survival model with a 28-day cycle length (which matches the typical treatment cycle length of the intervention and the comparators) and a lifetime time horizon. The model consisted of three health states: progression free (or pre-progression), progression (or post-progression), and death (Figure 2). The partitioned survival approach uses an "area under the curve" approach, where the number of patients in each health state at a given time is taken directly from survival curves fitted to the clinical data. This approach allows the survival of the comparator arms to be estimated using PFS and OS hazard ratios applied to the VEN+R survival curves. A half-cycle correction was applied in the base-case analysis.

The model assumes all patients enter the model in the pre-progression health state. Patients in the pre-progression health state, stay in that health state until disease progression. Transitions to the death state could occur from either the pre-progression or post-progression health state. Costs of disease management, utilities and risks of death all differ between the pre-progression and the post-progression health states. We note that many people with CLL may die of other causes.

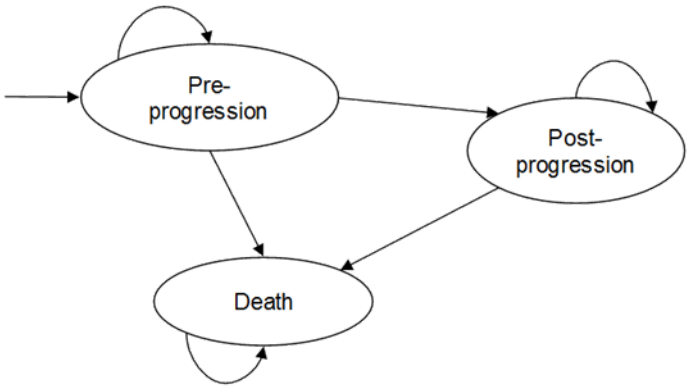


Figure 2: Model structure presented by the company

ERG summary

- The model takes a simple partitioned survival approach with three health states, and is consistent with other models built for patients with R/R CLL, and captures the two important clinical endpoints of OS and PFS.
- The cycle length of the model (28-days) is sufficiently short to capture changes over the relevant time interval.

5.2.3 Population

The population modelled in the company's base case analysis included:

- Patients with relapsed CLL - a CLL patient who previously achieved a CR or PR, but after a period of six or more months demonstrates evidence of disease progression;
- Patients with refractory CLL – a CLL patient who has progression within six months of the last anti-leukaemic therapy.

R/R CLL population is split into two further subgroups:

- patients with del(17p) and/or TP53 mutation
- patients with non-del(17p) and/or TP53 mutation.

Data for the base-case and the subgroup analyses were based on the MURANO study (a pooled dataset of the intervention and the control group). The study population was assumed by the

company to be reasonably similar to the UK population likely to receive treatment. However, out of the 389 patients recruited in the MURANO study, only 10 were from the UK (see section 4.2).

Data for ibrutinib arm came from RESONATE study[12] and data for the IDELA+R arm came from Study 116.[13]

Individuals in the modelled cohort had an average starting age of 64.18 years and 73.82% were male. An average body surface area (BSA) of 1.92m^2 was used to estimate the dosing of BR containing treatment regimens. The majority of patients (58.6%) in MURANO trial had at least one prior therapy, whereas 25.7% had at least two prior therapies. 26.96% of patients in the MURANO trial had del(17p) and/or TP53 mutation.

Information on patient characteristics for the subgroup analyses (i.e. del(17p)/TP53 and non-del(17p)/TP53) were not provided in the CS; furthermore, the ERG found that the mean values of the patient characteristics used in the base-case analysis were used in all subgroup analyses for the economic model.

ERG summary

- In the base-case analysis patients age and gender were taken from the overall trial population. However, the use of patient characteristics from only the European sites might result in more representative patients.
- The modelled population in all subgroup analyses were based on the characteristics of patients from the overall trial population, and not on the individual subgroups which were compared.

5.2.4 Interventions and comparators

In the company's base-case analysis, VEN+R is compared with ibrutinib or IDELA+R. Venetoclax is administered for a maximum of two years and rituximab is delivered for six cycles after completion of dose titration for venetoclax. The comparators ibrutinib and idelalisib are

administered until disease progression, and rituximab for the IDELA+R arm is administered for a total of six cycles.

The base-case economic model assumed that treatment effect with venetoclax lasted for a lifetime (approximately 30 years). But, the model also allowed for a treatment waning effect of 3 years after the discontinuation of venetoclax.

ERG summary

- The base-case analysis incorporates appropriate comparators relevant to the UK (ibrutinib or idelalisib+rituximab).

5.2.5 Perspective, time horizon and discounting

The perspective is as per NICE reference case, with benefits from a patient perspective and costs from an NHS and personal social services (PSS) perspective. A lifetime horizon is modelled (approximately 30 years). In the base-case, costs and benefits were discounted at an annual rate of 3.5%.

ERG summary

- The perspective, time horizon and discount rates chosen by the company all follow NICE recommendations, and are appropriate to the decision problem.

5.2.6 Treatment effectiveness and extrapolation

5.2.6.1 Survival Summary and Critique

In section B3.3.3, the company chose a partitioned survival model, and attempted to parameterise the observed OS and PFS curves from the MURANO trial in order to extrapolate and predict the long-term OS and PFS behaviour. Survival curves for the comparators were obtained by applying hazard ratios to the VEN+R curves.

5.2.6.2 VEN+R Time to Event Modelling

The company initially fitted models separately to each treatment arm's PFS and OS events, but, these extrapolations led to most curves predicting implausibly high OS for VEN+R, which exceeded the general population mortality. The ERG accept that these extrapolations were not suitable.

The company then chose to model PFS and OS jointly across both arms, assuming proportionality and the same parametric form between OS and PFS within and across both arms. They also included an interaction between treatment arm and endpoint (OS/PFS) allowing for the relationship between OS and PFS to be different across arms, and an interaction between del(17p)/TP53 status and endpoint, allowing del(17p) status to impact each outcome separately.

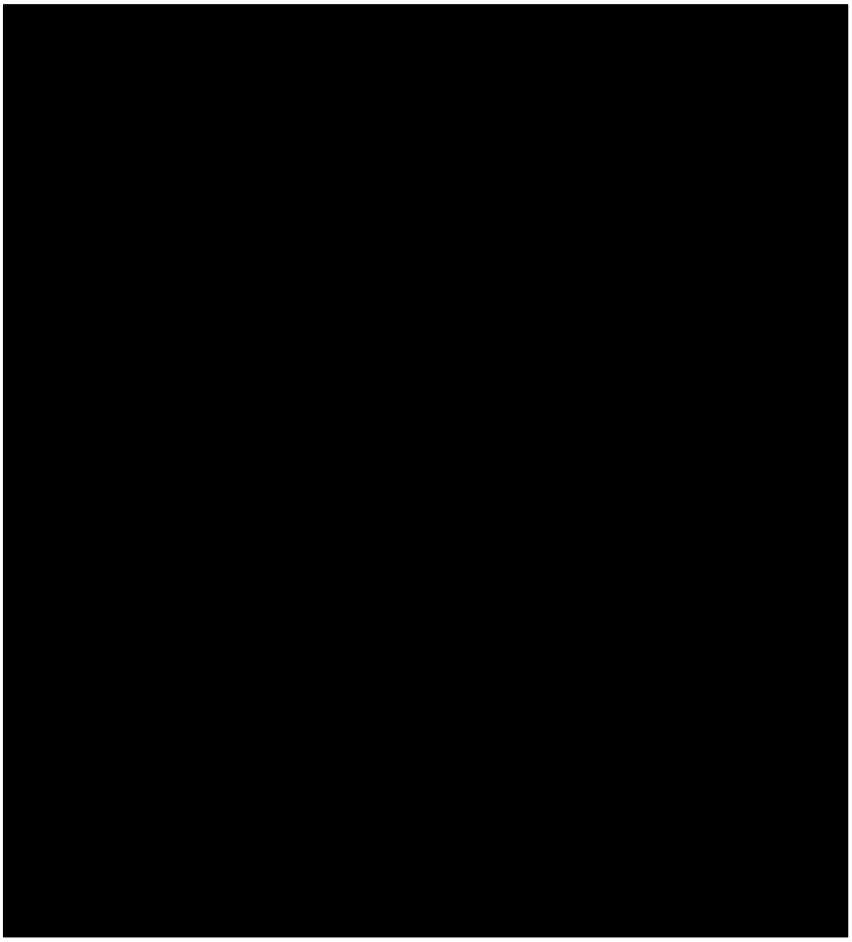
Whilst the model seems reasonable, the company does not provide any strong statistical evidence or description of the selection process of the model covariates, and so the ERG cannot comment on its robustness. It is unclear whether any other terms were considered for inclusion. The inclusion of the del(17p)/TP53 status and its interaction with the endpoint is questionable as the terms coefficients are not statistically significant in any of the parametric models presented in CS Table 35. Whilst the ERG appreciate that its inclusion enabled estimation of survival for the del(17p)/TP53 subgroup, it is not clear how helpful its inclusion is in the estimation of the full population model.

In order to verify the proportional hazards/survival-time assumption, the ERG requested additional evidence in the form of log-cumulative hazard plots. In general, proportionality did not appear strongly violated, though it is clear that the lines are not parallel in any of the plots, most evidently in the comparison of OS across both arms, shown in Figure 3. This means that whilst proportionality was violated, it was not done so to a statistically significant degree.

The ERG is also surprised at the decision of the company to include data from the BR arm of the trial when modelling OS and PFS, as this is not included as a comparator within the economic model. Thus, any HR referring to the relationship between the two arms should not have been estimated, and this means the VEN+R extrapolation of the immature OS data is influenced by the biologically different BR arm. However, the ERG does agree that the models produced without

such strong assumptions on proportionality produce implausible OS estimates (see clarification response B1). Nevertheless, the ERG is concerned that the company's decision to include covariates which may not significantly improve the model, combined with the inclusion of BR data may not result in a statistically robust analysis. This is supported by the resulting models fitted to the VEN+R OS data, shown in [redacted]. Here it is clear that the fitted curves do not reflect the observed data, which have resulted from the inclusion of BR data, in order to obtain plausible estimates. The ERG acknowledges the importance of an accurate extrapolation, but also feel that any modelling should also reflect observed data.

[redacted]3



[redacted]

The company assessed their jointly fitted parametric models through examination of their 20-year outcome predictions. Estimates were compared to the predictions made by five clinical experts, which is provided below in Table 16. These are in contrast to the estimates obtained from the company's jointly fitted survival curves in Table 17. It is apparent that despite the adjustments made by the company, a number of models still give implausible estimates. Exponential and Log-normal are too optimistic, and Gen-Gamma, 3-knot spline, and Gompertz are too pessimistic. However, Weibull, Log-logistic and Gamma all produce estimates of VEN+R 20 year OS that fall within the range of clinical expert opinions.

Table 16: Predictions from clinical experts on VEN+R long-term OS

Source	Expert Number	Prediction
Company	Clinical Expert 1	10% of patients alive at 20 years
Company	Clinical Expert 2	7% to 25% of patients alive at 20 years
Company	Clinical Expert 3	As high as 30% of patients alive at 20 years is reasonable, depending on the population
Company	Clinical Expert 4	Agreed with the more optimistic estimate (3)
Company	Clinical Expert 5	Agreed with the views of colleagues (1-5)
ERG	Clinical Expert 6	20-30% at 20 years
ERG	Clinical Expert 6	10-30% at 20 years (or matching the proportion of patients who are aged under 50 and achieved MRD negative status [17/194 patients])

Table 17: VEN+R OS predictions from company jointly fitted models

Outcomes			Exponential	Weibull	Gompertz	Log-logistic	Log-normal	Gamma	Gen gamma	3 knot spline
AIC			[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
VEN+R OS	Median (years)		[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	Survival %	2-year	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		5-year	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		10-year	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		20-year	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
BR OS	Median (years)		[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	Survival %	2-year	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		5-year	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		10-year	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		20-year	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

The company compared these estimates to three external data sources: 4-year follow-up from RESONATE,[29] fludarabine, cyclophosphamide, rituximab (FCR) data with 10-year follow-up[30] and 10-year registry data from the Haematological Malignancy Research Network (HMRN).[31]

The FCR data were from 284 patients, recruited in a phase II trial which began in December 1999. They had an observed 10-year OS of 23%, with extrapolations to 20 years performed by the

company ranging from 5% to 13%. This population was described by the company as healthier than that of MURANO due to being younger and in better general health. The HMRN data covered 2,723 patients diagnosed from September 2004 to August 2015, though it is unclear how many contributed to the second-line data considered in this analysis. The extrapolations ranged from 1% to 10% for 20-year OS, with the 8-year observed OS at approximately 18%.

However, the ERG do not believe these external studies are useful for predicting OS of VEN+R patients from MURANO. Firstly, the characteristics of the FCR study population show stark differences to the MURANO trial, as shown in Table 18. Large differences in age, Rai staging and presence of bulky disease. Baseline characteristics for the HMRN second-line population are not available, and so their similarity cannot be compared. Figure 5 demonstrates the large difference in observed OS between MURANO VEN+OS and the FCR data. Secondly, both FCR and HMRN began gathering data over 14 years ago, with major improvements in diagnosis and care increasing the heterogeneity to MURANO. Thirdly, it is unlikely that patients in these external studies received VEN+R, and so the ERG is unclear why they should be used to validate predictions made for VEN+R patients. The ERG believe these studies can only be used to exclude the Gompertz model (0% OS at 10 years), and not to distinguish between the plausibility of the remaining parametric models. Looking just at the observed periods from the external studies, both can be estimated to have 10-year OS in the region of 15%-25% once all participants data has been observed. However, comparing this to the 10-year predictions made from MURANO, it is clear that they are all much higher, ranging from 35.8% to 67%. The ERG are unsure why, given the apparent improvement of VEN+R at 10 years, why the company appear to predict that this benefit is lost at 20 years.

Table 18: Patient characteristics of VEN+R (MURANO) and FCR data

Effect modifier / prognostic characteristics	VEN+R	FCR
AGE >60	67.05%	45.77%
RAI III-IV	27.17%	45.77%
Bulky disease ≥ 5 cm	43.93%	7.14%
Beta Microglobulin > 3.5 mg/L	64.74%	59.93%
Prior therapy>1	44.51%	59.15%
CHROMOSOME 11Q DELETION	35.26%	12.75%
CRCL	26.59%	19.61%
Fludarabine Refractory	14.62%	19.01%
IGVH=Mutated	29.48%	31.40%
ECOG=1	45.56%	NR
Prior Purine Analog	80.47%	NR
Prior AntiCD20	73.96%	NR

NR = not reported

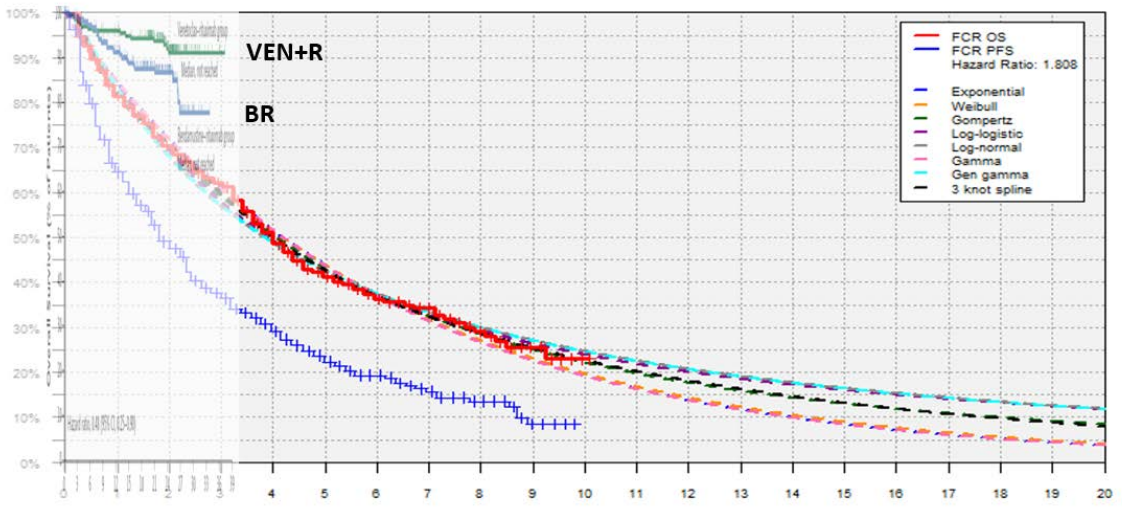


Figure 5: OS of MURANO overlaid onto Kaplan-Meier of FCR data

The ibrutinib data from RESONATE were also extrapolated, however with only four year's follow-up, there remained vast uncertainty in the extrapolations, with 20-year OS estimates ranging from 0% to 30%.

The Akaike information criteria (AIC) for the jointly fitted models were also provided by the company, however their relevance is limited as their calculation reflects the goodness of fit to the BR arm in addition to the VEN+R arm. As a result it is impossible to distinguish which is the best fitting model to the VEN+R arm alone.

The company state that the Weibull is their preferred parametric model for both OS and PFS and is used in their base-case analysis, supported by the external data. This results in 6.1 pre-progression life years (LY), and 4.7 post-progression LY, both undiscounted.

However, the ERG believe that the Weibull long-term predictions for OS may be too low, and expect to see a greater difference between the pre- and post- progression life years. The ERGs preference is to use the Gamma parametric model for OS, as it provides an OS more consistent with the above comparisons, which falls within the range of estimates from the clinical experts and has a lower AIC than the Log-logistic. In order to maintain the proportionality assumptions underlying the analysis, the ERG also chose a Gamma curve to model PFS.

Together, the Gamma curves slightly increases the ratio of PFS LY to post-progression survival (PPS) LY, versus the company's base-case. A comparison of the LY estimates, broken down into progression stage are shown in Table 19.

Table 19: Undiscounted LY estimates for VEN+R

	PFS	OS	PFS LY (% of total LY)	PPS LY (% of total LY)	Total LY
Company base-case	Weibull	Weibull	[redacted]	[redacted]	[redacted]
ERG preferred assumptions	Gamma	Gamma	[redacted]	[redacted]	[redacted]
ERG scenario	Log-logistic	Log-logistic	[redacted]	[redacted]	[redacted]

5.2.6.3 Ibrutinib

For their base-case, the company applied HR obtained from the MAIC to the parametric curves fitted to the VEN+R arm of MURANO. The ERG questioned this approach, given the critique of the MAIC in section 4.7 and section 4.8, the company's own admission that for the comparison to ibrutinib, the "HR estimates leads to a model dynamic which holds no face validity", and the ERG's own face validity checks (see section 5.2.14).

The company's economic model offered the option to model each comparator parametrically, based on curves fitted to the digitized IPD generated by the company. These curves were then adjusted depending on results of the MAIC analysis, to account for differences in baseline characteristics. Had the MAIC results been more clinically plausible, the ERG would have favoured this approach as it relaxed the assumptions of proportionality between the different treatments. However, as this approach is wholly reliant on the MAIC results, the ERG did not consider it an improvement on the HR based analysis.

It is the preference of the ERG to model the OS and PFS of ibrutinib using HR discussed in section 4.8, as this results in more plausible PPS estimates, as seen Table 20.

Table 20: Undiscounted LY estimates of ibrutinib

	PFS and OS Curves and HR	HR Source	PFS LY (% of total LY)	PPS LY (% of total LY)	Total LY
Company base-case	Weibull [redacted]	Company MAIC	4.64 (100%)	0.00 (0%)	4.64
ERG HR, company curves	Weibull [redacted]	ERG NMA	[redacted]	[redacted]	[redacted]
ERG preferred assumptions	Gamma [redacted]	ERG NMA	[redacted]	[redacted]	[redacted]

The survival curves for ibrutinib based on the company's and ERG's preferred assumptions are shown in [redacted], alongside the Kaplan-Meier (KM) data from RESONATE. The ERG believe that the company's assumptions result in a model that underestimates the effectiveness of ibrutinib in the MURANO population, given the similarity of the prediction to the observed OS in RESONATE, despite the difference in baseline populations.

6

5.2.6.4 Idelalisib + R

The ERG are concerned over the reliability of all the MAIC results, given the issues with the ibrutinib results. The ERG are reluctant to also use the resulting HRs for IDELA+R even though they appear plausible. However, the ERG were not able to find any comparisons of IDELA+R to BR and were unable to generate any alternative HRs. Hence, the ERG maintained the HRs estimated by the company, but apply them to the Gamma PFS and OS curves. The ERG also explored using the anchored MAIC results comparing IDELA+BR. The ERG acknowledges that IDELA+BR was not in the scope, and neither, the company or the ERG, found any evidence supporting any equivalence to IDELA+R. However, a comparison of the scenarios in Table 21 shows that it is the only scenario where PFS LY exceeds PPS LY, which the ERG expects in a disease such as CLL.

Table 21: Undiscounted LY estimates of IDELA+R

	PFS and OS Curves	HR Source	PFS LY (% of total LY)	PPS LY (% of total LY)	Total LY
Company base-case	Weibull	MAIC (IDELA+R)	1.80 (47%)	1.99 (53%)	3.79
ERG preferred assumptions	Gamma	MAIC (IDELA +R)	[redacted]	[redacted]	[redacted]
ERG alternative	Gamma	MAIC (IDELA +BR, adjusted)	[redacted]	[redacted]	[redacted]

ERG summary

- Company assume proportionality between OS, PFS and both arms of MURANO trial in order to gain plausible long-term estimates, suggesting data may be too immature to meaningfully extrapolate.
- Company prefer jointly fitted Weibull model for OS and PFS, due to similarity of 20-year OS prediction with external data and clinical expert opinion.
- ERG question the generalisability of the external data, and prefer jointly fitted gamma model, as this sustains some treatment benefit observed throughout the duration of the extrapolation.
- Company apply HR from their MAIC analysis to obtain predictions for ibrutinib and IDELA+R, despite some issues with the results.
- ERG prefer HR obtained from NMA, which result in a plausible balance of PFS and PPS LY for ibrutinib.

5.2.7 Mortality

General population background mortality was estimated using the latest UK life tables from the Office of National Statistics.[32]

5.2.8 Adverse events

The company outline their incorporation of AEs into the economic model in section B.3.3.5 of their submission. The CS state that only events of grade ≥3 that occurred in ≥5% of patients in any of the three main trials (MURANO, RESONATE and Study 116) were included. The ERG

believe this to be slightly inaccurate, as it appears that only AEs from the intervention arms of the three trials were considered (VEN+R; ibrutinib; IDELA+R) and that the AEs of comparator arms were not included. However, the ERG does not believe that this detracts from the relevance of the economic analysis presented by the company. The AEs included are shown below in Table 22 (adapted from CS Table 36), although 'infusion related reactions' were not reported in Table 36, they were included in the economic model. Across the majority of adverse event categories, the proportion of patients with an adverse event was generally higher in the intervention arm of the MURANO trial data than the intervention arms of the RESONATE and Study 116 trials. The only exception is pneumonia (6.19% for VEN+R and 6.67% for ibrutinib) and thrombocytopenia (6.17% for VEN+R, 5.65% for ibrutinib and 10.00% for IDELA+R). TLS was not included in the model as it did not meet the AE inclusion criteria. The ERG believe that AEs reported from MURANO may increase, as there were 78 patients receiving ongoing treatment at the point of data analysis.

Table 22: Adverse events used in the company's base-case analysis

AE	VEN+R	Ibrutinib	IDELA+R
N	194	195	110
Alanine aminotransferase (ALT)/Aspartate Transaminase (AST) elevation	1.55%	-	5.45%
Anaemia	10.82%	4.62%	5.45%
Autoimmune haemolytic anaemia	2.58%	-	-
Neutropenia	57.73%	16.41%	33.64%
Pneumonia	6.19%	6.67%	-
Thrombocytopenia	6.19%	5.64%	10.00%
Infusion Related Reaction	1.55%	-	-
Source	MURANO[11]	RESONATE[29]	Study 116[33]

The ERG note that 17.5% of patients in the venetoclax arm of MURANO experienced grade 3/4 infections or infestations, however, these were not included in the economic model with no explanation given.

The ERG also note that the frequencies of the AEs for VEN+R found in Table 36 (and **Error! Reference source not found.** above) do not all correspond to the frequencies found in CS Table 23. Whilst the incidence of events included in the economic model spans across grades 3-5, these frequencies are not presented within the clinical section of the CS or any other evidence found by the ERG. The ERG anticipate that the frequencies used in the company's base-case analysis are some combination of grade 3-4 AEs and SAEs, possibly with additional grade 5 events that have not been presented. The discrepancy of most concern is the frequency of pneumonia. The 6.19% incidence used for pneumonia is less than the pneumonia related SAEs (8.2%), and so the ERG believe this to be an error (see Table 23).

Table 23: Comparison of adverse event frequency across

AE	VEN+R CS Table 36 and company base-case (Grade 3-5)	VEN+R CS Table 23 Grade 3-4 AEs	VEN+R CS Table 23 SAEs
N	194	194	194
ALT/AST elevation	1.55%	-	-
Anaemia	10.82%	10.8%	1.5%
Autoimmune haemolytic anaemia	2.58%	-	-
Neutropenia	57.73%	57.7%	-
Pneumonia	6.19%	5.2%	8.2%
Thrombocytopenia	6.19%	5.7%	5.7%
Infusion Related Reaction	1.55%	1.5%	0.5%
Infection and Infestation	-	17.5%	-

The ERG have confirmed that the AE incidence for the ibrutinib and IDELA+R arms match the numbers reported in their corresponding main trial publications.[12, 13] However, values taken from the RESONATE trial, for ibrutinib, refer only to events of grade 3-4 and not grade 5. Hence, it is likely that AEs for ibrutinib may be slightly under-represented within the economic analysis.

Despite potential under-representation of AEs for ibrutinib and VEN+R, the ERG do not have any major concerns as these AEs are not a major driver of the cost-effectiveness analysis.

The ERG agrees with the CS approach in estimating QALY decrements associated with these adverse events, as a similar approach were used in previous appraisals for venetoclax monotherapy[8] and IDELA+R.[10] In brief, the estimates of the mean utility decrement and the mean duration associated with each adverse event were obtained from published sources including previous NICE technology appraisals and multiplied together to generate the required QALY decrement (CS Table 43). The ERG checked and verified that estimates of QALY decrements for adverse events reported in the CS are consistent with those reported in TA359.[10] No disutilities for TLS were included in the CS base-case model.

ERG summary

- General background mortality was taken from the latest UK lifetable estimates from Office of National Statistics.
- The company model included adverse events of grade ≥3 if they occurred in ≥5% of patients in any of the three main trials (MURANO, RESONATE and Study 116).
- TLS was not included in the model as an AE as it did not meet the AE inclusion criteria and therefore, no disutilities associated with TLS were included in the CS base-case model.
- 17.5% of patients in the venetoclax arm of MURANO experienced grade 3/4 infections or infestations, however, these were not included in the economic model.
- Estimates of QALY decrements for adverse events reported in the CS are consistent with those reported in TA359.

5.2.9 Health related quality of life

Health-related quality of life data were collected for MURANO trial participants using EQ-5D-3L; however, these health-state utility values derived from this data were not used to inform the economic model presented in the CS. The CS did not report the actual utility values derived from the MURANO trial data but explained that they were they were heavily skewed towards 1 or

"perfect health" and lacked face validity when compared to general UK adult population utility norms. Because of this, the CS did not use utility values derived from the MURANO trial to inform the subsequent economic model. However, upon clarification utility values were presented to the ERG; however, these utility values were not split by pre- or post-progression so were not used in any scenario analyses carried out by the ERG.

Instead, the CS used health state utility values from previous NICE technology appraisals of various technologies in CLL including venetoclax monotherapy (TA487)[8] and IDELA+R (TA359).[10] In these appraisals, a utility value of 0.748 was assigned to patients in pre-progression health state in the NICE committees most preferred base-case model [8, 10] and a mean utility of 0.600 for patients in the progressed health state, based on estimates reported in a published HTA report by Dretzke et al (2010)[34] and the subsequent appraisals of technologies in CLL. The company justified using these utility values on the grounds that they informed the committees' most preferred base-case model for venetoclax monotherapy[8] and IDELA+R.[10] Also, the post-progression health state utility value was based on data elicited directly from CLL patients rather than the general population, and was therefore considered the most robust utility value.[34]

In addition, to the health state utility values from the previous NICE technology appraisals mentioned above, the company conducted a systematic literature review to identify studies assessing health-related quality of life in R/R CLL. Detailed results of the review are presented in CS Appendix H with a summary presented in section B.3.4.3 of the CS. In total, 13 full-text articles were included in the final HRQoL review, two of which reported utility scores of 0.748 (CS Table 39) for the pre-progression health state.

The ERG agrees with the approach to health state utility estimation for the pre-progression and post-progression health states as used in the company's base-case model. The ERG notes the pre-progression utility of 0.748 and post-progression utility of 0.600 have been accepted in previous NICE committee deliberations as the most appropriate estimates of health utility in R/R CLL.[8, 10] and the ERG agrees that these utility values are the most appropriate for the patient population in the current appraisal of VEN+R as they are likely to be similar to the populations considered in TA487 and TA359.

The ERG further agrees with the CS reasons for not using utility values derived from the MURANO data in the economic model. It is noted that it highly unlikely that patients with R/R CLL have higher quality of life than the general adult population of a similar age and gender; hence, the health utility values derived from the MURANO data are likely to represent an over estimate of the actual HRQoL in patients with R/R CLL.

Uncertainty around these estimates of the mean pre-progression and post-progression utilities values and estimates of QALY decrements associated with adverse events was incorporated into the economic model by assuming that standard errors associated with each estimate equal to 10% of the mean.

Health-state utility values for pre-progression and post-progression health states and disutility associated with adverse events in the CS model were age-adjusted as recommended in NICE DSU TSD 18 to account for the increasing comorbidities with increasing age due to the resultant deterioration in quality of life in older aged cohorts.[24] Multiplicative adjustment factors were derived for age-groups between 60 and 85+ using pooled data from four consecutive health surveys for England (2003-2006) that reported health-stated utility values generated from the EQ-5D-3L health-state utility values.[35] The ERG agrees with the rationale for and the CS approach to adjusting for age-related utility deterioration.

ERG summary

- HRQoL data collected for MURANO trial participants using EQ-5D-3L lacked face validity to due to the health states utility values being higher than UK adult population norms.
- Health state utility values used in the economic model were taken from previous NICE technology appraisals in CLL.
- Patients in pre-progression health state were assigned a utility value of 0.748 and patients in post-progression health state were assigned a utility value of 0.60. Consistent with previous NICE committee decisions as the most appropriate estimates of health utility in R/R CLL patients.

- Health-state utilities and disutility associated with adverse events in the CS model were age-adjusted as recommended in NICE DSU TSD 18.

5.2.10 Resources and costs

5.2.10.1 Intervention and comparator costs

Tables 46 and 47 of the CS reproduced below for completeness summarises the CS approach to treatment regimen dosing and cost calculations for VEN+R and the comparator interventions (see Table 24 and Table 25). The costs for VEN+R for each cycle (28-days) in the CS were obtained from the BNF. Daily dose for venetoclax was 20 mg/day for week 1, 50 mg/day for week 2, 100 mg/day for week 3, 200 mg/day for week 4, and 400 mg/day for week 5 and beyond, up to a maximum treatment duration of 2 years. The model assumes intravenous (IV) rituximab is administered on day one of cycles 1 to 6 corresponding to a total of six doses of rituximab in first 6 months of treatment with VEN+R. Rituximab costs were estimated based on a dosing regimen of 375 mg/m^2 in day 1 of cycle 1 and 500 mg/m^2 in day 1 of cycles 2 to 6 and applying it to a body surface area of 1.92m^2 observed in the MURANO trial. There were no administration costs for venetoclax. Administration costs for rituximab were applied assuming 12 minutes of pharmacist time costing £9 per infusion based study by Millar et al.[36] and a 30:70 ratio between standard and rapid IV infusions for administration of rituximab containing treatment regimens. Unit costs for administration were obtained from the NHS Reference Costs 2016-17 and were £313.47 (HRG code SB15Z) for rituximab (IV standard) and £250.07 (HRG code SB12Z) for rituximab (IV Rapid).

Table 24: Drug acquisition costs (CS Table 46)

Drug	Pack size	Pack Cost	Per mg Cost	Source
Venetoclax	14 x 10 mg	£59.87	£0.43	BNF – 10, 50 and 100 mg tablets (AbbVie Ltd)
	7 x 50 mg	£149.67	£0.43	
	7 x 100 mg	£299.34	£0.43	
	14 x 100 mg	£598.68	£0.43	
	112 x 100 mg	£4,789.47	£0.43	

Rituximab (IV)	1 x 500 mg	£785.84	£1.57	BNF - Truxima 500 mg/50ml concentrate for solution for infusion vials (Napp Pharmaceuticals Ltd)
Rituximab (SC)	1 x 1,400 mg	£1,344.65	£0.96	NICE Evidence summary ESNM46 (2014)[37]
Ibrutinib	90 x 140 mg	£4,599.00	£0.37	BNF - Imbruvica 140 mg capsules (Janssen-Cilag Ltd)
Idelalisib	60 x 150 mg	£3,114.75	£0.35	BNF - Zydelig 150mg tablets (Gilead Sciences International Ltd)

Key: BNF, British National Formulary; IV, Intravenous; SC, Subcutaneous

Table 25: Treatment regimens (CS Table 47)

Regimen	Drug	Admin	Dosing schedule
VEN+R	Venetoclax	Oral	Daily dose, 20 mg week 1, 50 mg week 2, 100 mg week 3, 200 mg week 4, 400 mg week 5 and beyond until disease progression or 2-year maximum treatment duration.
	Rituximab	IV	375 mg/m^2 D1 C1, 500 mg/m^2 D1 C2-C6 for a total of 6 doses.
Ibrutinib	Ibrutinib	Oral	Daily dose of 420 mg until disease progression.
IDELA+R	Idelalisib	Oral	Daily dose of 300 mg until disease progression.
	Rituximab	IV	375 mg/m^2 D1 C1, 500 mg/m^2 D1 C2-C6 for a total of 6 doses.

The two comparator interventions of ibrutinib and IDELA+R were administered continuously until disease progression. Drug administration costs for ibrutinib was assumed to zero. Administration costs for IDELA+R were applied assuming treatment scheduling and costs similar to the assumptions applied in calculation of rituximab administration costs in the VEN+R (see Table 24 and Table 25).

No drug wastage costs were included in the model.

The ERG identified an error in the way intervention costs for VEN+R were applied in the CS economic model (See CS Table 49). The CS had applied the cost of rituximab in the first 6 cycles corresponding to approximately the first 6 months of treatment with VEN+R. The ERG believed the costs of rituximab should have been included in cycles 2 to 7 of the model because the dose-

titration schedule involves venetoclax monotherapy only in the first 4 weeks of treatment (corresponding to the first cycle of the model). The first dose of rituximab is given in week 5 (cycle 2 of the model) upon completion of venetoclax dose titration followed by 5 further doses of rituximab at the beginning of each cycle.[11] The ERG believe that the impact of this error in the CS model will be minimal because the error affects only the times at which rituximab costs were added in the model and not the total number of rituximab doses in the costing model. The ERG asked the company for clarification on this, please see section 5.3 for more detail.

5.2.10.2 Other health state costs

Other healthcare costs considered in the CS base-case economic model included the costs for TLS prophylaxis, other adverse events, 'routine care and monitoring' including hospital visits, investigations and procedures undertaken during a CLL patient's treatment pathway and the cost of terminal care.

TLS costs

The CS presented costs for TLS prophylaxis which were based on an algorithm along with its associated resource usage and costs in Tables 50 and 51 of the CS and in Appendix N. First, TLS was categorised into lower and greater risk groups based on the tumour mass and absolute lymphocyte count. So patients with lymph node diameter ≤5 cm and ALC <25 x 10^9/L indicates a low risk and all other patients are of a greater risk. Next, the high risk group is subdivided into two groups according to CRCL cut-off at 80 ml/min. The algorithm placed 18.06% of the MURANO trial population in the low risk group, 32.2% in the greater risk (CRCL≥80) group and 49.74% in the greater risk (CRCL<80) group (CS Table 50). Based on this algorithm, the cost of TLS prophylaxis applied in each cycle of the CS model were £1,430 for the low risk group, £2,016.54 for the greater risk (CRCL≥80) and £2,146.81 for the greater risk (CRCL<80).

The ERG notes a similar algorithm was used to derive TLS prophylaxis costs in TA487 (see Table 26). However, the estimated TLS costs were much higher in TA487 compared to the current submission (£1,808 for lower risk group, £2,235 greater risk group with CRCL≥80 and £2,334 for the greater risk group with CRCL<80). The ERG considered scenarios using the alternative higher estimates of TLS prophylaxis costs in its exploratory analyses.

Table 26: TLS prophylaxis costs by risk stratification

Submission	Lower risk	Greater risk	
		CRCL≥80	CRCL<80
Current submission	£1,430.40	£2,016.54	£2,146.81
TA487	£1,808	£2,235	£2,334

Key: ALC, absolute lymphocyte count; CRCL, creatinine clearance

Costs of routine care

The routine care costs take into account costs for the visits and procedures which occur during a CLL patient's treatment pathway. The resources and frequency usage were based on a previous NICE submission[9] and expert opinion which were detailed in CS Table 52. Resource use items in the economic model included: full blood counts, lactate dehydrogenase (LDH) tests, chest x-rays, bone marrow exams, haematologist visits, inpatient non-surgical medical stay, and blood and platelet transfusions. Unit costs were estimated based on NHS reference costs 2016/17.[38]

Pre-progression per cycle cost was estimated to be £27.12 and the post-progression per cycle cost was estimated to be £431.14. Table 27 presents the cost estimates associated with routine care from the CS alongside the routine care costs reported in TA487 (CS Table 69 of TA487).[39] The ERG noted that the pre-progression costs of £27.12 per cycle were substantially lower than the pre-progression estimate of £269.94 per cycle used in TA487 (see Table 27 below). The ERG was unable to find out what the key driver for this difference in the pre-progression routine care costs was, but notes that TA487 estimates also included costs for lymphocyte count, inpatient non-surgical medical stays, and nurse home visits that were not included in the pre-progression routine care costs calculations reported in the current submission. However, the CS indicated that feedback from clinician experts suggests the pre-progression health state resource use does not normally involved inpatient non-surgical medical visits and nurse home visits which may have an effect on reducing routine care costs in the pre-progression health state. The ERG considered scenarios using the alternative higher estimates of routine care costs in TA487 in its exploratory analyses.

Table 27: Routine care costs for patients with R/R CLL

Resource/procedure	CS model Table 53 (2017 prices)		TA487 - CS Table 69 (2016 prices)	
	Annual pre-progression frequency	Annual post-progression frequency	Annual pre-progression frequency	Annual post-progression frequency
Full blood count	4	8	4	4
LDH test	2	0	2	0
Lymphocyte count[1]	-	-	3.5	0
Chest x-ray	0	2	2	0
Bone marrow exam	0	1	1	0
Haematologist visit	2	6	4.5	4.9
Inpatient non-surgical medical stays	0	4	2	1
Nurse home visit[1]	-	-	3	4
Full blood transfusion	0	11	2	2
Platelet infusion[1]	-	-	0	0
Total annual cost	£353.78	£5,624.03	£3,509.17	£2,517.32
Per cycle cost	**£27.12**	**£431.14**	**£269.94**	**£193.64**

Other adverse events

The CS presented costs for adverse events in Table 54 (replicated below in Table 28); the majority of unit costs were obtained from NHS reference costs 2016/2017.[38] Adverse event costs associated with ALT/AST elevation were assumed to be zero based on previous NICE submission.[40] Costs used in NICE TA429[9] are shown in the second column in Table 28, which the ERG have explored using in a scenario analysis. Adverse events were applied only to the first cycle of the economic model for simplicity and there was a lack of information on when the AEs occurred for the comparators in the CS economic model.

Table 28: Summary of costs of AEs used in the economic model

AE	Costs used in company base-case	Costs used in NICE TA429	Costs used in ERG scenario analysis for VEN+R
ALT/AST elevation	£ 0.00	-	£0.00
Anaemia	£ 1,170.78	£ 3,042.17	£ 3,042.17
Autoimmune haemolytic anaemia	£ 1,170.78	-	£ 1,170.78
Neutropenia	£ 119.49	£ 2,386.17	£ 2,386.17
Pneumonia	£ 6,149.58	£ 2,733.21	£ 2,733.21
Thrombocytopenia	£ 621.34	£ 2,191.65	£ 2,191.65
Infusion Related Reaction	£ 401.07	-	£ 401.07

Terminal care costs

Terminal care costs were included in the economic model and applied to all patients who died. Cost estimates were based on a published study of end of life care for solid tumour cancer patients by Round et al (2015)[41] and were presented in CS Table 55. The specific cost used was guided by the TA429 appraisal.[9] The CS noted that clinical experts advising on the ibrutinib submission process suggested that the costs of terminal care would be similar between solid tumour and haematology patients. The total cost for terminal care per patient was £6,601.23 (inflated to 2016-17 prices).

ERG summary

- Drug dosing schedules and costs were provided by the company.
- No drug wastage costs were included in the model.
- A two-year stopping rule was applied when calculating intervention costs for VEN+R, whereas treatment with ibrutinib and IDELA+R continued until disease progression.
- Uncertainty exists around the sources used to estimate adverse event costs in the economic model. For this reason, the ERG have performed scenario analyses using estimates for adverse events from other sources identified in the literature.

5.2.11 Cost effectiveness results

5.2.11.1 Base-case analysis

The CS base-case analysis used PFS and OS hazard ratios from the unanchored MAIC, applying a 2-year maximum treatment duration to the VEN+R when estimating treatment costs, and assigning health-state utility values of 0.748 and 0.600 for the pre-progression and post-progression health states respectively.

The unanchored MAIC analysis that informed the base-case analysis generated a PFS HR of [redacted] and an OS HR of [redacted] for VEN+R vs. ibrutinib. For the comparison with IDELA+R, the corresponding HRs were [redacted] for PFS and [redacted] for OS, respectively. The CS noted that applying these HRs in the comparison with ibrutinib leads to PFS exceeding OS for ibrutinib which is impossible and lacks face-validity. Thus, in the CS base-case model, the PFS is restricted to being equal or lower than OS, resulting in zero post-progression period for ibrutinib.

The CS explained that this lacks face validity in the base-case model predictions of ibrutinib survival due to "*predominantly a consequence of the large uncertainty margins surrounding the MAIC estimates"*. However, the ERG notes that although the unanchored MAIC HRs had wide 95% CIs, this would not translate into uncertainty in the final cost-effectiveness estimates, because the MAIC estimates suggests VEN+R significantly improved OS compared to ibrutinib or IDELA+R (note OS estimates are the key drivers of cost-effectiveness in the CS sensitivity analyses).

The base-case model also accounted for disutility and costs associated with adverse events. The cost of TLS prophylaxis for patients on VEN+R are included in the model, but disutility associated with TLS was not taken into account. The CS base-case cost-effectiveness results for adults with R/R CLL who had at least received one prior therapy, with costs and QALYs discounted at 3.5% per annum over the 30-year time horizon are summarised in Table 29.

Table 29: Base-case discounted results, whole population (CS Tables 61 and 62)

Technologies	Total Costs, £	Total QALYs	Incremental Costs, £	Incremental QALYs	ICER vs. baseline (£/QALY)	Pairwise ICER vs. VEN+R (£/QALY)
No discount applied to VEN+R						
IDELA+R	[redacted]	2.307	[redacted]	-	[redacted]	[redacted]
VEN+R	[redacted]	5.666	[redacted]	3.358	[redacted]	[redacted]
Ibrutinib	[redacted]	3.067	[redacted]	-0.759	[redacted]	[redacted]
[redacted]**** applied to VEN+R***						
IDELA+R	[redacted]	2.307	-	-	-	£2,625
VEN+R	[redacted]	5.666	-7.003	-3.358	£2,625	-
Ibrutinib	[redacted]	3.067	-0.851	-0.759	£194,048	Dominated

* At net price ([redacted] applied to venetoclax)

For the adults with R/R CLL using list prices, the CS deterministic base-case showed that on average ibrutinib was the most expensive of the three interventions, but VEN+R generated more QALYs than ibrutinib or IDELA+R.

For the comparison with ibrutinib using the list price, the CS deterministic base-case showed VEN+R was [redacted] cheaper and also generated [redacted] more QALYs than ibrutinib. For the comparison with IDELA+R, VEN+R was more expensive, but generated more QALYs. Thus, the CS deterministic base-case analysis showed that VEN+R [redacted] ibrutinib; when VEN+R was compared with IDELA+R it generated an incremental cost-effectiveness ratio (ICER) of [redacted] per QALY gained.

The CS presented a deterministic base-case analysis in which a [redacted] is applied to the list price of venetoclax in the VEN+R regimen (CS Table 62). These cost-effectiveness results were very similar to those based on list price with VEN+R dominating ibrutinib; and generating an ICER of £2,625 per QALY gained when comparing VEN+R with IDELA+R (see Table 29).

5.2.11.2 Probabilistic base-case analysis

The CS presented probabilistic base-case analysis incorporating uncertainty in the model inputs. This allows for the probability that each intervention is the most cost-effective strategy to be calculated. The CS probabilistic base-case results produced similar results to the deterministic analysis with VEN+R dominating ibrutinib, and when compared with IDELA+R generating a probabilistic mean ICER of [redacted] per QALY gained.

Cost-effectiveness planes from the CS clarification response for the probabilistic base-case analysis using both the list and the net prices ([redacted] for VEN+R) are presented in Figure 7 and Figure 8. When VEN+R is compared with ibrutinib the majority of iterations fall in the south-east quadrant; whereas, when VEN+R is compared with IDELA+R the majority of the iterations fall in the north-east quadrant.

[redacted]

8

The cost-effectiveness acceptability curves (CEACs) from the CS probabilistic base-case analysis using both the list and the net prices (for VEN+R) are presented in Figure 9 and Figure 10. These show that the probability that VEN+R is cost-effective compared ibrutinib, and when VEN+R is compared with IDELA+R at a willingness-to-pay threshold of £20,000 per QALY the probability was close to based on the list price analysis and over when based on the net price analysis.

9

5.2.12 Sensitivity analyses

5.2.12.1 Deterministic sensitivity analysis

The CS conducted one-way sensitivity analysis (OWSA) to identify key model drivers and important sources of uncertainty by varying or substituting alternative values of parameter inputs one at a time. In each of these analyses, the central estimate of each base-case parameter was replaced with lower and higher estimates that correspond to the lower and upper 95% CIs of parameter inputs. Tornado plots showing the first six-parameters associated with the greatest uncertainty on cost-effectiveness results on the net monetary benefit scale are presented in Figure 11 and Figure 12 for the list price comparisons with ibrutinib and IDELA+R. The plots suggests that the OS and PFS hazard ratios and the VEN+R joint model parameters had the greatest impact on incremental costs and incremental QALYs (and hence the incremental net monetary benefit) in the comparison with ibrutinib.

10

Key: BR, bendamustine+rituximab; HR, hazard ratio; OS, overall survival; PFS, progression free survival; TLS, tumour lysis syndrome; VEN+R, venetoclax+rituximab

11

Key: BR, bendamustine+rituximab; HR, hazard ratio; OS, overall survival; PFS, progression free survival; TLS, tumour lysis syndrome; VEN+R, venetoclax+rituximab

For the comparison with ibrutinib, the ERG believes the CS deterministic OWSA that used the upper and lower 95% CI estimates of the HR for OS from the unanchored MAIC were not that informative and potentially misleading because the unanchored MAIC analysis does not adequately capture the uncertainty in overall survival estimates for VEN+R vs. ibrutinib. This is because the MAIC results suggested VEN+R significantly improved OS compared with ibrutinib by considerable margin (i.e. crudely, OS HRs translate into almost [redacted] in the hazard/risk of death for VEN+R compared with ibrutinib on average, 95% CIs ranging from [redacted] to [redacted] reduction in risk of death). The CS claims that due to the immaturity of the MURANO trial data (see CS section B.2), estimates of HRs for OS based on the MAIC analysis are highly uncertain, but the ERG does not believe the 95% CIs around the OS HR for VEN+R vs. ibrutinib reflected any degree of uncertainty (when used to inform a deterministic cost-effectiveness model) because the HRs suggested that VEN+R significantly improved OS compared with ibrutinib. This combined with the 2-year maximum treatment duration for VEN+R implies VEN+R will continue to dominate ibrutinib when using the estimate of OS HRs from the unanchored MAIC analysis. The ERG believes a more informative OWSA exploring uncertainty with the OS benefit for VEN+R compared with ibrutinib will have been to use the OS HRs from the anchored MAIC

analysis that compared VEN+R to ibrutinib under the assumption that the relative efficacy of VEN+R vs. ibrutinib+BR can be extended to VEN+R vs. ibrutinib.[20] The OS HRs from the anchored MAIC suggested [redacted]. This confidence interval crosses 1 and hence reflects a greater degree of uncertainty in the comparison with ibrutinib.

5.2.12.2 Scenario analyses

The CS presented extensive scenario analyses to test the robustness of the model structure and assumptions (see CS Tables 68 and 69). In all, a total of 51 analyses were conducted for R/R CLL using both list and net prices (with the net price analysis applying a [redacted] discount to the cost of VEN in the VEN+R regimen). The CS found the model predictions were generally robust with VEN+R continuing to dominate ibrutinib in the majority of the scenario analyses undertaken. The only exception to this trend reported was when the analyses are restricted to shorter time horizons (1-year and 2-year) when using the list price. When comparing ibrutinib with VEN+R, the ICERs were [redacted] and [redacted] per QALY gained based on shorter 2-year and 1-year time horizons, respectively.

5.2.13 Subgroup analyses

The CS presented cost-effectiveness results for subgroup of R/R CLL patients with (i) del(17p) and/or TP53 mutation and (ii) without del(17p) and/or TP53 mutation. The CS explained that del(17p) and TP53 mutation are known to negatively affect a patient's prognosis, thus patients with this mutation would generally have a lower survival than the whole R/R CLL population and those patients who do not have this deletion or mutation (see CS Figures 43 to Figure 45).

The net effect of this is that average time to treatment (ToT) for the treatment regimens are considerably shorter for patients with del(17p)/TP53 as shown in Table 30 (combining data displayed in CS Table 58, Table 70 and Table 75).

Table 30: Average time on treatment

Treatment	Average time on treatment (Mean years)		
	Whole R/R CLL population	del(17p)/TP53 subgroup	Non-del(17p)/TP53 subgroup
VEN+R	1.859	1.823	1.871
Ibrutinib	4.661	3.965	4.880
IDELA+R	1.833	1.535	1.957

Cost-effectiveness results for the subgroup of patients with and without del(17p)/TP53 from the CS are presented in Table 31 and Table 32 respectively, and they are in in line with company's base-case results.

Table 31: Base-case results (del(17p)/TP53) (CS Table 73 and 74)

Technologies	Total Costs, £	Total QALYs	Incremental Costs, £	Incremental QALYs	ICER vs. baseline (£/QALY)	Pairwise ICER VS. VEN+R (£/QALY)
No discount applied to VEN+R						
IDELA+R	[redacted]	2.045	[redacted]	-	[redacted]	[redacted]
VEN + R	[redacted]	5.132	[redacted]	-3.087	[redacted]	[redacted]
Ibrutinib	[redacted]	2.726	[redacted]	-0.681	[redacted]	[redacted]
[redacted] ***applied to VEN+R***						
IDELA+R	[redacted]	2.045	-	-	-	£6,013
VEN + R	[redacted]	5.132	-£18,558	-3.087	£6,013	-
Ibrutinib	[redacted]	2.726	-£127,669	-0.681	£187,556	Dominated

Table 32: Base-case results (non-del(17p)/TP53) (CS Table 78 and 79)

Technologies	Total Costs, £	Total QALYs	Incremental Costs, £	Incremental QALYs	ICER vs. baseline (£/QALY)	Pairwise ICER VS. VEN+R (£/QALY)
No discount applied to VEN+R						
IDELA+R	[redacted]	2.411	[redacted]	-	[redacted]	[redacted]
VEN + R	[redacted]	5.869	[redacted]	-3.458	[redacted]	[redacted]
Ibrutinib	[redacted]	3.193	[redacted]	-0.782	[redacted]	[redacted]
[redacted] ***applied to VEN+R***						
IDELA+R	[redacted]	2.411	-	-	-	£1,333
VEN + R	[redacted]	5.869	-£4,608	-3.458	£1,333	-
Ibrutinib	[redacted]	3.193	-£152,538	-0.782	£194,985	Dominated

5.2.14 Model validation and face validity check

5.2.14.1 Company's work

The CS reported a number model validation and face-validity checks following the structured format described in the Assessment of the Validation Status of Health-Economic decision models (AdViSHE) checklist.[42] This included:

- Assessment of face-validity and conceptual model structure check by a number of health economists and academics (including [redacted]) experienced in critique of economic models in CLL submitted for reimbursement decisions by NICE.
- Cross validating the model by comparing the model structure and outcomes to that of other economic models in CLL (including models that informed previous TAs). Cross validation of model results of existing models were not explicitly conducted.
- Scenario analyses incorporating alternative input data were used to cross-validate model inputs (section B.3.8.3 of CS).
- Reported quality checks and tests (and tests results) carried by senior economic modeller of the excel model (CS Table 81).

5.2.14.2 ERG's face validity check

As indicated in section 4.7, the ERG has found that the OS HR estimate for the VEN+R versus ibrutinib comparison, which was obtained from the MAIC comparison, was not plausible given its magnitude and the implausible relationship between PFS and OS HRs. The use of this HR in the cost-effectiveness evaluation to compare VEN+R vs ibrutinib led to an estimated life expectancy of 10.78 years for VEN+R and 4.63 years for ibrutinib. Below the ERG has further demonstrated that the estimated life expectancy with ibrutinib derived from the company's model is pessimistic.

First, the ERG has attempted to compare the predictions made by the company to the previous appraisal of ibrutinib[9]. However, the estimated LYs reported in the publicly available committee papers were redacted, and only the incremental LYs were visible, as shown in Table 33. The ERG of NICE TA429 commented that whilst the indirect comparisons of ibrutinib suggested it was clinically superior to its comparators, there remained significant uncertainty over the magnitude of the benefit.

Table 33: Incremental LYG estimates of ibrutinib

Comparators	Incremental life year gain NICE TA429	Estimates from company's base-case
Ibrutinib vs Ofatumumab	3.47 (Head to Head Trial)	-
Ibrutinib vs Idelalisib+Rituximab	2.60 (Bucher ITC)	0.85
Ibrutinib vs Bendamustine+Rituximab	4.79 (MAIC)	-
Source	Taken from Table 8 from company comments to ACD1 of NICE TA429 [9]. Unclear if discounting has been applied.	Obtained from economic model, undiscounted. (4.635 - 3.785)

It is clear that LY of ibrutinib estimated from the company's base-case compared to that of VEN+R analysis is far more pessimistic than in TA429. Despite the LY being withheld from TA429, it is apparent that the estimate is very likely to exceed 5 years due to the estimated incremental difference against BR. However, the estimate of undiscounted LYs in the company's base-case analysis for ibrutinib was just 4.6 years (information extracted from the company's economic model and CS Table 61).

When using an OS HR of 0.48 for the comparator treatments, as estimated in the clinical section for the relative efficacy of BR to VEN+R (Seymour et al. NEJM[11]), the undiscounted estimated LY is [redacted] years. Adding the 4.79 years incremental LYs estimated for ibrutinib in TA429 would imply a total LY of [redacted] years for ibrutinib. This contrasts greatly with the 4.63 years reported in the CS.

Second, the ERG has undertaken further analysis by digitizing published OS KM graph[29] from the RESONATE study. Using DigitizeIt v2.2.3[43] software, IPD was generated, replicating the ibrutinib population. This IPD was then modelled parametrically using Stata 15 [26], and the mean survival calculated accordingly.

As reported by the company in the modelling of the MURANO data, the more flexible parametric models predicted a decreasing hazard rate over time, which is known not to reflect the true long nature of the disease. As a result, an exponential model provided the most plausible estimate, which assumed a constant hazard over time and so it is possible that this approach produces a slightly optimistic estimate of the ibrutinib life years, however it is similar to the extrapolation used in the appraisal of ibrutinib, where the company initially opted for a log-normal curve followed by an exponential tail.

The resulting life expectancy from the second method is [redacted] years for ibrutinib.

Table 34 shows the LY estimates using our two methods described above compared to that obtained from the company's economic model which used the MAIC-estimated OS HR. The ERG also compared the median OS predicted by the company's base-case, to the ERG's preferred HR under the company's assumptions, and to the ERG's reconstructed IPD (Table 35). Both of the ERG's approaches estimate a much higher median OS than the company's base-case. Our methods demonstrate that the company's estimate of 4.635 years is pessimistic. In addition, published 3-year OS data is available from the RESONATE study[20], with 74% of patients on ibrutinib alive. The company's base-case model predicts that only [redacted] of patients will be alive at 3 years, further demonstrating the poor representation of ibrutinib in the company's model, despite the fact that the baseline characteristics of the trials suggest that MURANO population is healthier.

In section 5.3, we will show that the use of OS HR derived from the indirect treatment comparison undertaken by the ERG (section 4.8) leads to much more plausible life expectancy for ibrutinib which matches with the estimates reported in Table 33.

Table 34: A comparison of the ibrutinib LY estimates

	Company's model (derived using MAIC OS HR, undiscounted)	ERG's method 1: using incremental difference from TA429 of ibrutinib and BR, applied to estimate of BR LYG from MURANO (unclear if discounting is applied)	ERG's method 2: using reconstructed IPD from RESONATE+ extrapolation (undiscounted)
Ibrutinib life expectancy estimate	[redacted]4.635	[redacted]	[redacted]

Table 35: Comparison of median OS for VEN+R and ibrutinib.

Treatment	Scenario	Assumptions	Median OS
VEN+R	Company Base-case	Weibull Curve	[redacted]
Ibrutinib	Company Base-case	MAIC HR applied to VEN+R Weibull survival	[redacted]
Ibrutinib	ERG NMA HR for Ibrutinib	ERG HR applied to VEN+R Weibull survival	[redacted]
Ibrutinib	ERG IPD reconstruction	Exponential Curve	[redacted]

5.3 Exploratory and sensitivity analyses undertaken by the ERG

The ERG undertook extensive exploratory analyses to assess the effect of varying model assumptions and parameter inputs on the cost-effectiveness results. As stated in section 5.2.10, the ERG identified an error in the model that meant the cost of rituximab was applied to the VEN+R regimen in first 6 cycles of the model (corresponding to approximately the first 6 months of treatment). The ERG believed the costs of rituximab should be included in the cycles 2 to 7 of the model because the dose-titration schedule involves venetoclax monotherapy only in the first 4 weeks of treatment (corresponding to the first cycle of the model). The ERG asked the company to clarify whether the costs of rituximab were included in cycle 7 of the model for VEN+R regimen. In response, the company confirmed that *"the cost of rituximab is not included in the 7th cycle onwards for the total treatment costs of VEN+R and idelalisib+R. The dosing regimen of rituximab used in the model is 375 mg/m2 administered on day 1 of cycle 1 and 500 mg/m2 on day 1 of cycles 2-6 for a total of 6 cycles"*. The ERG believes the CS approach to calculation of rituximab costs is not correct for the reasons given, but we don't believe that the total costs or the

ICER would change much should a correction be made. The company did not provide an economic model with this correction in the clarification response.

In response to further clarifications raised by the ERG about rituximab in the VEN+R arm after the clarification process was completed, the company stated that "*Rituximab is administered after completion of the dose titration period of venetoclax. However, the model simplifies such that venetoclax (dose titration) and rituximab start on the same day; structural changes would be required to bring this into alignment with the MURANO protocol and would have minimal impact on results*."

However, the company did note the following: "*upon investigating the dose titration assumption in the model more closely, it has come to our attention that an error has occurred regarding the time on venetoclax treatment. According to the MURANO protocol, venetoclax dosing at 400mg should be given to progressive disease or 2 years, from start of combination therapy. However, in the model, the dose titration period has been captured in this 2-year duration, and one cycle of venetoclax at 400mg has been erroneously excluded……"* The company then provided guidance for correcting the error so that modelling of VEN+R dosing regimen closely matches that specified in the MURANO trial. The correction involves including an additional cycle for venetoclax (i.e. treatment cycle changes from 24 to 25) and also additional week of venetoclax (400 mg per day) in the titration period. The company provided updated base-case results generated from the corrected models for the R/R CLL population which showed that ICER for VEN+R vs ibrutinib remains [redacted] while the ICER for VEN+R vs IDELA+R increases by [redacted] to [redacted] per QALY gained. The company also stated that "the corrections made also influence the budget impact however, the impact is moderate."

Cost-effectiveness results generated using the company's base-case parameters applied to the corrected model are presented in Table 36. When using the list prices, the results suggest VEN+R remained [redacted] compared with ibrutinib, whilst the ICER for VEN+R compared with IDELA+R increased from [redacted] per QALY gained in the original CS base-case model to [redacted] per QALY gained in the corrected model. Using the net price after applying a [redacted] discount for VEN+R, the ICER increased from £2,625 per QALY gained in the original CS base-case model to £3,492 per QALY gained when compared with IDELA+R.

Table 36: CS base–case corrected model: CS base-case discounted results after ERG applied the corrections to the dosing regimen and treatment costs for VEN+R for R/R CLL population

Technologies	Total Costs, £	Total QALYs	Incremental Costs, £	Incremental QALYs	ICER vs. VEN+R (£/QALY)
No discount applied to VEN+R					
VEN+R	[redacted]	5.666	[redacted]	-	
Ibrutinib	[redacted]	3.067	[redacted]	2.599	[redacted]
IDELA+R	[redacted]	2.307	[redacted]	3.358	[redacted]
[redacted] ***applied to VEN+R***					
VEN+R	[redacted]	5.666	-	-	
Ibrutinib	[redacted]	3.067	-£135,650	2.599	Dominated
IDELA+R	[redacted]	2.307	£11,726	3.358	£3,492

The ERG exploratory analyses reported below are based on the corrected model.

The CS base-case model was informed by HRs derived from adjusted MAIC analyses. Thus, the ERG believes the modelled population should therefore have been the competitor trial population when using the MAIC estimates and not from the MURANO trial. For the comparison with ibrutinib, this would involves adjusting the mean age, % male and % with del(17p)/TP53 mutation from 64.2 years, 73.8% and 29.96% observed in MURANO trial to 66.5 years, 68.0% and 32.3% in the RESONATE cohort, respectively. Similarly for the comparison with IDELA+R, the modelled population should be adjusted to median age of 71 years, 73.8% male and 43.64% with del(17p)/TP53 mutation reflecting the distribution of these characteristics in Study 116. Implementing these changes have very minimal impact on the cost-effectiveness estimates with VEN+R continuing to [redacted] ibrutinib in both list and net price comparisons (Table 37). For the comparison with IDELA+R, the ICER increased by [redacted] (list price) and by [redacted] (net price) per QALY gained (Table 38).

Table 37: CS base–case corrected model: changed modelled population to the RESONATE in the comparison with ibrutinib (R/R CLL population)

Technologies	Total Costs, £	Total QALYs	Incremental Costs, £	Incremental QALYs	ICER vs. VEN+R (£/QALY)
No discount applied to VEN+R					
VEN+R	[redacted]	5.55	[redacted]	-	
Ibrutinib	[redacted]	3.017	[redacted]	2.533	[redacted]
[redacted] ***applied to VEN+R***					
VEN+R	[redacted]	5.55	-	-	
Ibrutinib	[redacted]	3.017	-£133,765	2.533	Dominated

Table 38: CS base–case corrected model: changed modelled population to Study 116 cohorts in the comparison with IDELA+R (R/R CLL population)

Technologies	Total Costs, £	Total QALYs	Incremental Costs, £	Incremental QALYs	ICER vs. VEN+R (£/QALY)
No discount applied to VEN+R					
VEN+R	[redacted]	5.24	[redacted]	-	
IDELA+R	[redacted]	2.156	[redacted]	3.084	[redacted]
[redacted] ***applied to VEN+R***					
VEN+R	[redacted]	5.24	£102,033	-	
IDELA+R	[redacted]	2.156	£13,815	3.084	£4,480

5.3.1 Uncertainty around the OS hazard ratio in the comparison with ibrutinib

For the comparison with ibrutinib, the company provided anchored MAIC estimates in the CS as sensitivity analyses under the assumption that ibrutinib single-agent has equivalent efficacy to ibrutinib+BR based on the results of Hillmen et al (2015).[20] Under this assumption, anchored MAIC analyses could be conducted assuming that relative efficacy of VEN+R vs. ibrutinib+BR could be extended to VEN+R vs. ibrutinib single-agent (see CS section B.2.9.5).

The OS hazard ratio from the anchored MAIC was [redacted] The results presented in Table 39 to Table 41 suggests that:

- Applying the mean and lower 95% CI estimate of the OS HR had minimal impact on the ICER with VEN+R continuing to [redacted] ibrutinib based on both the list and net price comparisons (Table 39 and Table 40).
- Applying the higher 95% CI estimate of the OS HR (i.e. [redacted]) generated an incremental cost of [redacted] (list price analysis), [redacted] (net price analysis) and incremental QALYs of [redacted] for VEN+R vs. ibrutinib (Table 41). This suggests that VEN+R is cheaper but also generated fewer QALYs on average than ibrutinib. The ICER was [redacted] (list price) and [redacted] (net price analysis) [redacted] for VEN+R compared with ibrutinib.

Table 39: CS base–case corrected model: used OS HR from company's anchored MAIC (adjusted) analysis (R/R CLL population)

Technologies	Total Costs, £	Total QALYs	Incremental Costs, £	Incremental QALYs	Pairwise ICER (£/QALY)
No discount applied to VEN+R					
Ibrutinib	[redacted]	4.191	[redacted]	-	[redacted]
VEN + R	[redacted]	5.666	[redacted]	1.475	[redacted]
[redacted] ***applied to VEN+R***					
Ibrutinib	[redacted]	4.191			
VEN + R	[redacted]	5.666	-£149,447	1.475	Dominated

Table 40: CS base–case corrected model: used lower 95% CI estimate of the OS HR from company's anchored MAIC (adjusted) analysis (R/R CLL population)

Technologies	Total Costs, £	Total QALYs	Incremental Costs, £	Incremental QALYs	Pairwise ICER (£/QALY)
No discount applied to VEN+R					
Ibrutinib	[redacted]	2.397	[redacted]	-	[redacted]
VEN + R	[redacted]	5.666	[redacted]	3.269	[redacted]
[redacted] ***applied to VEN+R***					

Ibrutinib	[redacted]	2.397			
VEN + R	[redacted]	5.666	-£84,647	3.269	Dominated

Table 41: CS base–case corrected model: used upper 95% CI estimate of the OS HR from company's anchored MAIC (adjusted) analysis (R/R CLL population)

Technologies	Total Costs, £	Total QALYs	Incremental Costs, £	Incremental QALYs	Pairwise ICER (£/QALY)
No discount applied to VEN+R					
Ibrutinib	[redacted]	6.546	[redacted]	-	[redacted]
VEN + R	[redacted]	5.666	[redacted]	-0.88	[redacted]
[redacted] ***applied to VEN+R***					
Ibrutinib	[redacted]	6.546			
VEN + R	[redacted]	5.666	-£172,056	-0.88	£195,564

5.3.2 Uncertainty around the OS hazard ratio in the comparison with IDELA+R

The ERG has conducted exploratory analyses similar to those carried out for the ibrutinib comparison to investigate uncertainties around the OS HR for VEN+R vs IDELA+R. Using data that the company provided in response to ERG clarification questions (see point 3, section A9), the company explained that HRs for OS and PFS for VEN+R vs IDELA+BR were based on anchored MAIC analysis and these were not presented in the original CS because there is no published evidence to suggest IDELA+R and IDELA+BR have similar efficacy. Nevertheless, the company provided adjusted anchored MAIC estimates suggesting that VEN+R is associated with PFS HR of [redacted] based on the IRC definition of PFS and OS HR of [redacted] compared with IDELA+BR. The ERG was satisfied with the company's response and appreciates the effort undertaken for the extra set of analysis.

The ERG agrees with the company that HRs generated from the anchored MAIC analysis that compared VEN+R vs. IDELA+BR were not appropriate for the decision problem. The ERG conducted its own literature review but was unable to identify studies that would allow an indirect comparison between VEN+R vs IDELA+R. In the absence of reliable comparative evidence, the ERG conducted a sensitivity analyses to test the impact of assuming similar effect for VEN+R and IDELA+R by setting the HR for OS for VEN+R vs. IDELA+R to 1 (Table 42). Under this

assumption, VEN+R was more costly but generated more QALYs than IDELA+R generating an ICER of [redacted] per QALY gained in the list price analysis. For the net price analysis, VEN+R was cheaper and generated more QALYs than IDELA+R, therefore dominated IDELA+R.

Table 42: CS base–case corrected model: assumed an OS HR of 1 for VEN+R vs. IDELA+R (R/R CLL population)

Technologies	Total Costs, £	Total QALYs	Incremental Costs, £	Incremental QALYs	Pairwise ICER (£/QALY)
No discount applied to VEN+R					
IDELA+R	[redacted]	5.154	[redacted]	-	[redacted]
VEN + R	[redacted]	5.666	[redacted]	0.512	[redacted]
[redacted] ***applied to VEN+R***					
IDELA+R	[redacted]	5.154			
VEN + R	[redacted]	5.666	-£14,944	0.512	Dominated

5.3.3 ERG preferred method of estimating the hazard ratio for VEN+R vs. ibrutinib

The company's adjusted unanchored MAIC analysis produced an OS HR of [redacted] for VEN+R vs. ibrutinib, suggesting a [redacted]% risk reduction in OS with VEN+R compared with ibrutinib. As already stated, the ERG believed this HR is highly uncertain.

Therefore, the ERG conducted an indirect comparison using a fixed-effect NMA to compare survival outcomes for VEN+R vs. ibrutinib (see section 4.8), using these new HRs from the indirect comparison the ERG applied this to corrected base-case model. As seen in Table 43, the CS base-case corrected ICER changed from VEN+R dominating ibrutinib, to an ICER of [redacted] (list price) and £790,988 (net price) per QALY lost (i.e. VEN+R was cheaper but also generated on average 0.354 fewer QALYs compared with ibrutinib).

Table 43: CS base–case corrected model: used central estimate of PFS and OS HR for VEN+R vs. ibrutinib from ERG's indirect comparison analysis (R/R CLL population)

Technologies	Total Costs, £	Total QALYs	Incremental Costs, £	Incremental QALYs	Pairwise ICER (£/QALY)
No discount applied to VEN+R					
Ibrutinib	[redacted]	6.019	[redacted]	-	[redacted]
VEN + R	[redacted]	5.666	[redacted]	-0.354	[redacted]
[redacted] ***applied to VEN+R***					
Ibrutinib	[redacted]	6.019			
VEN + R	[redacted]	5.666	-£279,766	-0.354	£790,988

Using the lower and upper 95% CI estimate of HRs generated from the ERG's indirect comparison in OWSA suggested that the cost-effectiveness results were most sensitive to the HR for OS with ICERs ranging from VEN+R [redacted] ibrutinib using the lower 95% CI estimate to VEN+R being comparatively cheaper, but also generating fewer QALYs than ibrutinib using the upper 95% CI estimate for OS (see Table 51 for further sensitivity analyses results).

5.3.4 Further exploratory analyses undertaken by ERG

The ERG considered the company's approach to parameterisation and long-term extrapolation of the OS and PFS curves for VEN+R and the comparators (see section 5.2.6). The ERG conducted a series of exploratory analysis based on the corrected model to investigate the impact of assuming alternative parametric modelling of PFS and OS. The results suggest changing the parametric modelling from joint-Weibull to joint-Gamma survival curves for both OS and PFS (Table 44) had minimal impact on the ICER with VEN+R continuing the [redacted] ibrutinib in both list and net price comparisons. For the comparison with IDELA+R, the ICER decreased from [redacted] to [redacted] per QALY gained based on list price analysis and from [redacted] to £2,903 per QALY gained based on net price analysis (Table 44).

Table 44: CS base–case corrected model: changed PFS and OS parametric curves from joint-Weibull to joint-Gamma: VEN+R vs ibrutinib (R/R CLL population)

Technologies	Total Costs, £	Total QALYs	Incremental Costs, £	Incremental QALYs	ICER vs. VEN+R (£/QALY)
No discount applied to VEN+R					
VEN+R	[redacted]	6.04	[redacted]	-	
Ibrutinib	[redacted]	3.157	[redacted]	2.884	[redacted]
IDELA+R	[redacted]	2.351	[redacted]	3.69	[redacted]
[redacted] ***applied to VEN+R***					
VEN+R	[redacted]	6.04	-	-	
Ibrutinib	[redacted]	3.157	-£142,716	2.884	Dominated
IDELA+R	[redacted]	2.351	£10,711	3.69	£2,903

The ERG also tried further analyses, for example, where we choose joint-Gamma for PFS and joint-Weibull for OS (or vice versa), but this had minimal impact on the ICER, whereby VEN+R continued to [redacted] ibrutinib and the ICERs ranged between [redacted] and [redacted] per QALY gained for VEN+R compared with IDELA+R (see Table 51 and Table 52).

The ERG considered scenarios using the alternative higher estimates of routine care costs and TLS prophylaxis costs based on the figures in TA487 and adverse events costs based on Figures reported in TA439 (see Section 5.2.10.2). Implementing all these changes together had minimal impact on the ICER with VEN+R continuing to [redacted] ibrutinib (Table 45). For the comparison with IDELA+R, the ICER increased from the CS corrected base-case value of [redacted] to [redacted] per QALY gained based on list price and from [redacted] to £5,694 per QALY gained based on the net price (Table 45).

Table 45: CS base–case corrected model: changed TLS prophylaxis, adverse events costs and routine care costs (R/R CLL population)

Technologies	Total Costs, £	Total QALYs	Incremental Costs, £	Incremental QALYs	ICER vs. VEN+R (£/QALY)
No discount applied to VEN+R					
VEN+R	[redacted]	5.666	[redacted]	-	
Ibrutinib	[redacted]	3.157	[redacted]	2.884	[redacted]
IDELA+R	[redacted]	2.307	[redacted]	3.358	[redacted]
[redacted] ***applied to VEN+R***					
VEN+R	[redacted]	5.666	-	-	
Ibrutinib	[redacted]	3.157	-£142,716	2.884	Dominated
IDELA+R	[redacted]	2.307	£19,123	3.358	£5,694

5.3.5 ERGs preferred base-case model

5.3.5.1 ERGs preferred base-case for the ibrutinib comparison

The ERG's preferred base-case model for the ibrutinib comparison involves making the following assumptions and changes to the CS corrected base-case model:

- Changing the parametric survival curves from joint-Weibull to joint-Gamma for both PFS and OS
- Changing the unanchored MAIC PFS and OS HRs to ERGs indirect comparison using estimates of PFS and OS for ibrutinib vs BR reported in Hillmen (2015)[20] and for VEN+R vs BR based on the MURANO data.

The ERGs preferred base-case for the comparison with ibrutinib is presented in Table 46.

Table 46: ERG preferred base–case corrected model for the comparison with ibrutinib (R/R CLL population)

Technologies	Total Costs, £	Total QALYs	Incremental Costs, £	Incremental QALYs	ICER vs. VEN+R (£/QALY)
No discount applied to VEN+R					
VEN+R	[redacted]	6.04	[redacted]	-	
Ibrutinib	[redacted]	6.431	[redacted]	-0.39	[redacted]
[redacted] ***applied to VEN+R***					
VEN+R	[redacted]	6.04	-	-	
Ibrutinib	[redacted]	6.431	-£322,979	-0.39	£827,252

The results in Table 46 suggest VEN+R is [redacted] (list prices) and -£322,979 (net prices) cheaper than ibrutinib, but also generated 0.39 fewer discounted QALYs on average. The corresponding ICERs were [redacted] and £827,252 per QALY lost for VEN+R compared with ibrutinib based on list and net price comparisons, respectively. The ERG preferred base-case corrected model thus produced similar estimate of incremental costs as the CS base-case corrected model but differed in the direction of incremental QALYs generated. The ERG probabilistic base-case results (not presented) produced similar ICERs as the deterministic analyses. The probability that VEN+R is cost-effective compared with ibrutinib at £20,000 per QALY is close to [redacted] in both the list and net price comparisons.

The ERG applied its preferred base-case model to the populations with and without del(17p)/TP53 mutation for the ibrutinib comparison. The results of these analyses were similar to the ERGs preferred base-case results with VEN+R being cheaper but also generating fewer QALYs compared with ibrutinib in both list and net prices comparison (Table 47 and Table 48).

Table 47: ERG preferred base–case corrected model (del(17p)/TP53 mutation) for the comparison with ibrutinib

Technologies	Total Costs, £	Total QALYs	Incremental Costs, £	Incremental QALYs	ICER vs. VEN+R (£/QALY)
No discount applied to VEN+R					
VEN+R	[redacted]	5.494	[redacted]	-	
Ibrutinib	[redacted]	5.87	[redacted]	-0.376	[redacted]
[redacted] ***applied to VEN+R***					
VEN+R	[redacted]	5.494	-	-	
Ibrutinib	[redacted]	5.87	-£269,728	-0.376	£718,043

Table 48: ERG preferred base–case corrected model (nondel(17p)/TP53 mutation)) for the comparison with ibrutinib

No discount applied to VEN+R					
VEN+R	[redacted]	6.245	[redacted]	-	
Ibrutinib	[redacted]	6.638	[redacted]	-0.393	[redacted]
[redacted] ***applied to VEN+R***					
VEN+R	[redacted]	6.245	-	-	
Ibrutinib	[redacted]	6.638	-£343,718	-0.393	£873,858

5.3.5.2 ERGs preferred base-case model with a waning effect for the ibrutinib comparison

Due to the two-year treatment course of venetoclax for patients receiving VEN+R, the ERG believe it is plausible that the effects of VEN+R on OS and PFS may wane over time, thus increasing the hazard. Waning effects are often implemented through a steady or sudden increase in a hazard rate of the intervention relative to the hazard rate of one of the comparators. However in this appraisal, a waning effect was incorporated into the model through a percentage increase in the predicted hazards for VEN+R, after 5 years, i.e. increasing the hazard of VEN+R relative to

itself. The ERG are unclear why the company chose this approach and they did not instead chose to wane the hazard of VEN+R to either BR, external data or to one of the main comparators.

The ERG are also unsure over the justification for the fixed 5-year implementation point and would have preferred greater flexibility over the beginning of the waning effect. The company also chose to explore the effect of various hazard increases applied simultaneously to PFS and OS (20%, 50% and 100%), again the percentages were chosen arbitrarily. Without any suitable reference or anchor treatment, the ERG found it difficult to establish a range of plausible values for their own sensitivity analysis, and so applied the company's hazard increases onto the ERG base-case assumptions, and also considered scenarios with 10% and 70% hazard increases.

Table 49: ERG preferred base–case model with waning effect applied to PFS and OS estimates for VEN+R in the comparison with ibrutinib (R/R CLL population)

ERG exploration	Total costs VEN+R	Total LYs VEN+R	Total QALYs VEN+R	Total costs Ibrutinib	Total LYs Ibrutinib	Total QALYs Ibrutinib	Incremental costs	Incremental LYs	Incremental QALYs	ICER (LYs)	ICER (QALYs)
No discount applied to VEN+R											
ERG preferred base-case model	[redacted]	8.976	6.04	[redacted]	9.302	6.431	[redacted]	-0.326	-0.39	[redacted]	[redacted]
Applied 10%	[redacted]	8.647	5.832	[redacted]	9.302	6.431	[redacted]	-0.655	-0.599	[redacted]	[redacted]
Applied 20%	[redacted]	8.351	5.645	[redacted]	9.302	6.431	[redacted]	-0.951	-0.786	[redacted]	[redacted]
Applied 50%	[redacted]	7.621	5.182	[redacted]	9.302	6.431	[redacted]	-1.682	-1.249	[redacted]	[redacted]
Applied 70%	[redacted]	7.234	4.937	[redacted]	9.302	6.431	[redacted]	-2.068	-1.494	[redacted]	[redacted]
Applied 100%	[redacted]	6.761	4.636	[redacted]	9.302	6.431	[redacted]	-2.541	-1.795	[redacted]	[redacted]
[redacted] ***applied to VEN+R***											
ERG preferred base-case model	[redacted]	8.976	6.04	[redacted]	9.302	6.431	-£322,979	-0.326	-0.39	£989,832	£827,252
Applied 10%	[redacted]	8.647	5.832	[redacted]	9.302	6.431	-£323,590	-0.655	-0.599	£493,888	£540,430
Applied 20%	[redacted]	8.351	5.645	[redacted]	9.302	6.431	-£324,179	-0.951	-0.786	£340,860	£412,418
Applied 50%	[redacted]	7.621	5.182	[redacted]	9.302	6.431	-£325,781	-1.682	-1.249	£193,730	£260,920
Applied 70%	[redacted]	7.234	4.937	[redacted]	9.302	6.431	-£326,700	-2.068	-1.494	£157,946	£218,679
Applied 100%	[redacted]	6.761	4.636	[redacted]	9.302	6.431	-£327,878	-2.541	-1.795	£129,028	£182,682

The ERG's exploratory analyses in which it applied different rates of waning effect to the venetoclax had the effect of reducing survival outcomes and hence, the total number of life-years lived, total costs and total QALYs for VEN+R. For the list price comparisons, the ICER for VEN+R versus ibrutinib decreased from [redacted] per QALY lost in the ERG's preferred base-case model to between [redacted] per QALY lost for a 10% waning effect and [redacted] per QALY lost with 100% waning effect (Table 49). A similar downward trend in the ICER with an increasing waning effect is observed in the net price comparisons when a [redacted] discount is applied to venetoclax (Table 49).

5.3.5.3 ERGs preferred base-case for the IDELA+R comparison

The ERG was unable to conduct a preferred base-case analysis for the comparison with IDELA+R because no robust estimates of relative efficacy between VEN+R vs. IDELA+R was available. The ERG does not have confidence in the robustness of HRs generated from the company's unanchored MAIC analysis. The ERG conducted a scoping review of the literature but was unable to find relevant information that could be used to estimate the relative effectiveness of VEN+R vs. IDELA+R.

5.4 Conclusions of the cost effectiveness section

The CS presented an economic model that evaluated the cost-effectiveness of VEN+R vs. ibrutinib and IDELA+R as treatment options for adult patients with R/R CLL. The MURANO trial was the main source of clinical effectiveness evidence.

The company extrapolated OS and PFS using a jointly fitted Weibull model to both arms and to both outcomes of the MURANO trial, with strong assumptions of proportionality necessary to obtain plausible OS predictions for VEN+R. The ERG preferred to use a gamma model, which is more consistent with the external data considered by the company, but have concerns of the immaturity of the OS data and its suitability for extrapolation.

The two main drivers of cost-effectiveness versus ibrutinib were the 2-year fixed treatment duration for VEN+R and the HR for OS. The latter was estimated from an unanchored MAIC that

the company had performed. However, the ERG had major reservations about the robustness of the MAIC analyses and the HRs generated from it. For example, the magnitude of the OS benefit that VEN+R had over ibrutinib in the unanchored MAIC would suggest that ibrutinib had worse OS than BR. This the ERG felt is highly implausible based on published evidence on relative efficacy of ibrutinib versus BR.

The ERG identified an error in the calculation of intervention costs for VEN+R which the company corrected upon clarification.

In the company's original and corrected base-case models, VEN+R [redacted] ibrutinib and generated ICERs between [redacted] and [redacted] per QALY gained in the comparison with IDELA+R.

The ERG's preferred base-case model that used HRs from an indirect comparison performed by the ERG suggested that VEN+R was associated with lower costs and lower QALYs compared with ibrutinib with ICERs between [redacted] and £827,252 per QALY lost in the analyses that used list and net prices for VEN+R, respectively.

The ERG was unable to conduct a preferred base-case analysis for the comparison with IDELA+R due to lack of clinical effectiveness evidence for VEN+R vs. IDELA+R.

Further exploratory analyses conducted by the ERG suggested the ICERs were robust to different model inputs and very similar for patients with and without the del(17p)/ TP53 mutation.

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

Alterations to the base-case assumptions were made by the ERG as identified in Chapter 5. Further exploratory analyses undertaken by the ERG to test the robustness of the CS base-case assumptions and parameter inputs are in the Appendix. Results are presented in Table 51 for the comparison with ibrutinib and Table 52 for comparison with IDELA+R.

The impact on each change individually on the base-case analysis in comparison with ibrutinib is shown in Table 50.

Table 50: ERG re-estimation of cost-effectiveness

	ΔC	**ΔQALY**	**ΔC/QALY**	**Ratio[+]**
Comparison with ibrutinib – list price				
CS base-case corrected model	[redacted]	2.599	[redacted]	-
ERG models				
Changing the unanchored MAIC PFS and OS HRs to ERGs indirect comparison using estimates of PFS and OS for ibrutinib vs BR reported in Hillmen and for VEN+R vs BR based on the MURANO data	[redacted]	-0.354	[redacted]	[redacted]
ERG preferred base-case analysis	[redacted]	-0.39	[redacted]	[redacted]
Comparison with ibrutinib – net price				
CS base-case model	-£135,650	2.599	Dominated	-
ERG models				
Changing parametric survival curves from joint Weibull to joint-Gamma for both PFS and OS	-£142,716	2.884	Dominated	-

Changing the unanchored MAIC PFS and OS HRs to ERGs indirect comparison using estimates of PFS and OS for ibrutinib vs BR reported in Hillmen and for VEN+R vs BR based on the MURANO data	-£279,766	-0.354	£790,988	-
ERG preferred base-case analysis	-£322,979	-0.39	£827,252	-

+ The ERG have not calculated the ratio

The ERG was unable to conduct a preferred base-case model for the comparison with IDELA+R because no robust estimates of relative efficacy between VEN+R vs. IDELA+R was available.

7 END OF LIFE

End of life considerations do not apply.

8 OVERALL CONCLUSION

8.1 Clinical effectiveness evidence

Although the absence of relevant direct evidence justified the company's decision to conduct a MAIC analysis of VEN+R versus single agent ibrutinib, and the methods used in matching trial populations have been previously validated, the ERG remains concerned about the imprecise estimates of the resulting treatment effect of VEN+R (confidence intervals of HRs for PFS and OS were wide) as well as the implausible HRs for OS. Additional work undertaken by the ERG indirectly comparing estimates of the treatment effect of VEN+R from the MURANO trial against single-agent ibrutinib from Hillmen and colleagues[20] supports the ERG's position.

8.2 Cost-effectiveness evidence

The ERG conducted extensive exploratory analyses to understand the key drivers of cost-effectiveness and to explore the full extent of uncertainty in the economic model results. Absolute lymphocyte count However, there remains a considerable degree of uncertainty associated with the final estimates of cost-effectiveness because the key parameter in the economic model, the hazard ratio for overall survival that measures the magnitude of treatment benefit for VEN+R versus the comparator interventions was estimated with high degree of uncertainty in both the company's submission and the ERG exploratory analyses.

9 REFERENCES

1. Swerdlow SH, Campo E, Harris N, Jaffe ES, Pileri SA, Stein H, *et al.*, editors. *WHO classification of tumours of haematopoietic and lymphoid tissues. Fourth Edition*. Lyon, France: International Agency for Research on Cancer, World Health Organization; 2008.

2. Oscier D, Dearden C, Eren E, Fegan C, Follows G, Hillmen P, *et al.* Guidelines on the diagnosis, investigation and management of chronic lymphocytic leukaemia. *British Journal of Haematology* 2012;**159**:541-64.

3. Rodrigues CA, Goncalves MV, Ikoma MR, Lorand-Metze I, Pereira AD, Farias DL, *et al.* Diagnosis and treatment of chronic lymphocytic leukemia: recommendations from the Brazilian Group of Chronic Lymphocytic Leukemia. *Rev Bras Hematol Hemoter* 2016;**38**:346-57. http://dx.doi.org/10.1016/j.bjhh.2016.07.004

4. Cancer Research UK. *Chronic lymphocytic leukaemia (CLL) incidence statistics*. URL: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/leukaemia-cll/incidence (Accessed 1 August 2018).

5. Shanafelt TD, Rabe KG, Kay NE, Zent CS, Jelinek DF, Reinalda MS, *et al.* Age at diagnosis and the utility of prognostic testing in patients with chronic lymphocytic leukemia. *Cancer* 2010;**116**:4777-87. http://dx.doi.org/10.1002/cncr.25292

6. Cancer Research UK. *Chronic lymphocytic leukemia (CLL) incidence by age*. URL: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/leukaemia-cll/incidence#heading-One (Accessed 1 July 2018).

7. Cancer Research UK. *Chronic lymphocytic leukaemia (CLL) survival statistics*. URL: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/leukaemia-cll/survival#heading-Zero (Accessed 1 August 2018).

8. National Institute for Health and Care Excellence. Venetoclax for treating chronic lymphocytic leukaemia. NICE technology appraisal guidance 487. 2017.

9. National Institute for Health and Care Excellence. Ibrutinib for previously treated chronic lymphocytic leukaemia and untreated chronic lymphocytic leukaemia with 17p deletion or TP53 mutation. NICE technology appraisal guidance 429. 2017.

10. National Institute for Health and Care Excellence. Idelalisib for treating chronic lymphocytic leukaemia. NICE technology appraisal guidance 359. 2015.

11. Seymour JF, Kipps TJ, Eichhorst B, Hillmen P, D'Rozario J, Assouline S, *et al.* Venetoclax-Rituximab in Relapsed or Refractory Chronic Lymphocytic Leukemia. *The New England Journal of Medicine* 2018;**378**:1107-20. http://dx.doi.org/10.1056/NEJMoa1713976

12. Byrd JC, Brown JR, O'Brien S, Barrientos JC, Kay NE, Reddy NM, *et al.* Ibrutinib versus ofatumumab in previously treated chronic lymphoid leukemia. *The New England Journal of Medicine* 2014;**371**:213-23. http://dx.doi.org/10.1056/NEJMoa1400376

13. Furman RR, Sharman JP, Coutre SE, Cheson BD, Pagel JM, Hillmen P, *et al.* Idelalisib and rituximab in relapsed chronic lymphocytic leukemia. *The New England Journal of Medicine* 2014;**370**:997-1007. http://dx.doi.org/10.1056/NEJMoa1315226

14. Beauchemin C, Johnston JB, Lapierre ME, Aissa F, Lachaine J. Relationship between progression-free survival and overall survival in chronic lymphocytic leukemia: a literature-based analysis. *Curr Oncol* 2015;**22**:e148-56. http://dx.doi.org/10.3747/co.22.2119

15. Signorovitch JE, Swallow E, Kantor E, Wang X, Klimovsky J, Haas T, *et al.* Everolimus and sunitinib for advanced pancreatic neuroendocrine tumors: a matching-adjusted indirect comparison. *Experimental hematology & oncology* 2013;**2**:32. http://dx.doi.org/10.1186/2162-3619-2-32

16. Hallek M, Cheson BD, Catovsky D, Caligaris-Cappio F, Dighiero G, Döhner H, *et al.* Guidelines for the diagnosis and treatment of chronic lymphocytic leukemia: a report from the International Workshop on Chronic Lymphocytic Leukemia updating the National Cancer Institute-Working Group 1996 guidelines. *Blood* 2008;**111**:5446-56. http://dx.doi.org/10.1182/blood-2007-06-093906

17. Hallek M, Cheson BD, Catovsky D, Caligaris-Cappio F, Dighiero G, Döhner H, *et al.* Guidelines for diagnosis, indications for treatment, response assessment and supportive management of chronic lymphocytic leukemia. *Blood* 2018;**131**:2745-60. http://dx.doi.org/10.1182/blood-2017-09-806398

18. Rawstron A, Cohen D, De Tute R, McParland L, Collett L, Phillips D, *et al.* Bone marrow is more sensitive than peripheral blood for detection of MRD in cll and provides a more reliable prediction of outcome across different treatments. Paper presented at: 20th Congress of the European Hematology Association; Vienna, Austria, 11-14 June 2015.

19. Brander DM, Choi MY, Roberts AW, Ma S, Lash LL, Verdugo M, *et al.* Detailed safety analysis of venetoclax combined with rituximab in patients with relapsed/refractory chronic lymphocytic leukemia. *Blood Conference: 58th Annual Meeting of the American Society of Hematology, ASH* 2016;**128**:2033.

20. Hillmen P, Fraser G, Jones J, Rule S, O'Brien S, Dilhuydy MS, *et al.* Comparing Single-Agent Ibrutinib, Bendamustine Plus Rituximab (BR) and Ibrutinib Plus BR in Patients with Previously Treated Chronic Lymphocytic Leukemia/Small Lymphocytic Lymphoma (CLL/SLL): An Indirect Comparison of the RESONATE and HELIOS Trials. *Blood* 2015;**126**:2944.

21. Chanan-Khan A, Cramer P, Demirkan F, Fraser G, Silva RS, Grosicki S, *et al.* Ibrutinib combined with bendamustine and rituximab compared with placebo, bendamustine, and rituximab for previously treated chronic lymphocytic leukaemia or small lymphocytic lymphoma (HELIOS): a randomised, double-blind, phase 3 study. *The Lancet Oncology* 2016;**17**:200-11. http://dx.doi.org/10.1016/s1470-2045(15)00465-9

22. Betts KA, Mittal M, Song J, Skup M, Joshi A. OP0115 Relative Efficacy of Adalimumab versus Secukinumab in Active Ankylosing Spondylitis: A Matching-Adjusted Indirect Comparison. *Annals of the Rheumatic Diseases* 2016;**75**:98-9. http://dx.doi.org/10.1136/annrheumdis-2016-eular.2754

23. Maksymowych W, Strand V, Baeten D, Nash P, Thom H, Cure S, *et al.* OP0114 Secukinumab for The Treatment of Ankylosing Spondylitis: Comparative Effectiveness Results versus Adalimumab Using A Matching-Adjusted Indirect Comparison. *Annals of the Rheumatic Diseases* 2016;**75**:98-. http://dx.doi.org/10.1136/annrheumdis-2016-eular.2050

24. Phillippo D, Ades aE, Dias S, Palmer S, Abrams K, Welton N. NICE DSU technical support document 18: methods for population-adjusted indirect comparisons in submissions to NICE. 2016.

25. Hosmer DW, Lemeshow S, Sturdivant RX. *Applied logistic regression*. 3rd edn. Hoboken, New Jersey: Wiley; 2013.

26. *Stata*. Texas, USA: StataCorp LLC. URL: https://www.stata.com/ (Accessed 1 July 2018).

27. Drummond M, Jefferson T. Guidelines for authors and peer reviewers of economic submissions to the BMJ. The BMJ Economic Evaluation Working Party. *BMJ: British Medical Journal* 1996;**313**:275.

28. Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D, *et al.* Consolidated health economic evaluation reporting standards (CHEERS) statement. *Cost Effectiveness and Resource Allocation* 2013;**11**:6.

29. Byrd J, Hillmen P, O'Brien S, Barrientos JC, Reddy NM, Coutre S, *et al.* Long-term efficacy and safety with ibrutinib (ibr) in previously treated chronic lymphocytic leukemia (CLL): Up to four years follow-up of the RESONATE study. *J Clin Oncol* 2017;**35**:7510. http://dx.doi.org/10.1200/JCO.2017.35.15_suppl.7510

30. Badoux XC, Keating MJ, Wang X, O'Brien SM, Ferrajoli A, Faderl S, *et al.* Fludarabine, cyclophosphamide, and rituximab chemoimmunotherapy is highly effective treatment for relapsed patients with CLL. *Blood* 2011;**117**:3016-24. http://dx.doi.org/10.1182/blood-2010-08-304683

31. Haematological Malignancy Research Network (HMRN). *Clinical Management and Outcome in Chronic Lymphocytic Leukaemia. Version 2.1 [Draft]*. 2017. URL: https://www.hmrn.org/Download.aspx?target=document&id=296 (Accessed 1 July 2018).

32. Office for National Statistics. *National life tables: United Kingdom: 2014-2016*. 2017. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/lifeexpectancies/datasets/nationallifetablesunitedkingdomreferencetables (Accessed 1 August 2018).

33. Sharman JP, Coutre SE, Furman RR, Cheson BD, Pagel JM, Hillmen P, *et al.* Second Interim Analysis of a Phase 3 Study of Idelalisib (ZYDELIG®) Plus Rituximab (R) for Relapsed Chronic Lymphocytic Leukemia (CLL): Efficacy Analysis in Patient Subpopulations with Del(17p) and Other Adverse Prognostic Factors. *Blood* 2014;**124**:330.

34. Dretzke J, Barton P, Kaambwa B, Connock M, Uthman O, Bayliss S, *et al.* Rituximab for the treatment of relapsed/refractory chronic lymphocytic leukaemia. *Health Technol Assess* 2010;**14**:19-26.

35. Ara R, Brazier JE. Using health state utility values from the general population to approximate baselines in decision analytic models when condition-specific data are not available. *Value in Health: The Journal of the International Society for Pharmacoeconomics and Outcomes Research* 2011;**14**:539-45. http://dx.doi.org/10.1016/j.jval.2010.10.029

36. Millar DR, Corrie P, Hill M, Pulfer A. PCN74 a service evaluation to compare secondary care resource use between xelox and folfox-6 regimens in the treatment of metastatic colorectal cancer (MCRC) from a UK national health service (NHS) perspective. *Value in Health: The Journal of the International Society for Pharmacoeconomics and Outcomes Research* 2008;**11**:A483.

37. National Institute for Health and Care Excellence. *Non-Hodgkin's lymphoma: rituximab subcutaneous injection. NICE evidence summary (ESNM46)*. 2014. URL: https://www.nice.org.uk/advice/esnm46/chapter/Full-evidence-summary (Accessed 3 July 2018).

38. NHS Improvement. *Reference costs 2016/2017: National schedule of reference costs 1*. 2017. URL: https://improvement.nhs.uk/documents/1973/2_-_National_schedule_of_reference_costs_-_the_main_schedule.xlsx (Accessed 10 June 2018).

39. National Institute for Health and Care Excellence. *Venetoclax for treating chronic lymphocytic leukaemia. Technology appraisal guidance [TA487]: committee papers*. 2017. URL: https://www.nice.org.uk/guidance/ta487/documents/committee-papers (Accessed 1 August 2018).

40. National Institute for Health and Clinical Excellence. *Rituximab for the treatment of relapsed or refractory chronic lymphocytic leukaemia. NICE technology appraisal guidance 193*. 2010).

41. Round J, Jones L, Morris S. Estimating the cost of caring for people with cancer at the end of life: A modelling study. *Palliative medicine* 2015;**29**:899-907. http://dx.doi.org/10.1177/0269216315595203

42. 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**:349-61. http://dx.doi.org/10.1007/s40273-015-0327-2

43. Bormann I. *DigitizeIt*. Braunschweig, Germany: Bormisoft. URL: http://www.digitizeit.de/ (Accessed 1 July 2018).

10 APPENDIX

Table 51: Further exploratory analyses undertaken by the ERG for the comparison with ibrutinib

ERG exploration	List price comparisons			Net price comparisons			Cell changes in GEN SETTINGS sheet in economic model*
	Incremental costs	Incremental QALYs	ICER	Incremental costs	Incremental QALYs	ICER	
CS corrected base-case	[redacted]	[redacted]	[redacted]	-£135,650	2.599	Dominated	-
Changed mean age, % male and % del(17p) to values in RESONATE	[redacted]	[redacted]	[redacted]	-£133,765	2.533	Dominated	C8, C9 & C11
Changed routine care costs to £269.94 (pre-progression) and £193.64 (post-progression) figures in TA487	[redacted]	[redacted]	[redacted]	-£141,853	2.599	Dominated	G15 & H15 in CostCalcs sheet
Changed all AE costs to values used in TA439	[redacted]	[redacted]	[redacted]	-£134,524	2.599	Dominated	C34 to C39
Changed TLS prophylaxis costs to £1,808, £2,235 & £2,334 (TA487) for lower risk, greater risk (CRCL ≥80) & greater risk (CRCL<80) groups respectively	[redacted]	[redacted]	[redacted]	-£135,419	2.599	Dominated	M126, N126 & O126 in TLS prophylaxis sheet
Changed TLS prophylaxis and routine care costs to figures reported in TA487; and AE costs to the figures in NICE TA439	[redacted]	[redacted]	[redacted]	-£140,496	2.599	Dominated	C34 to C39; G15 & H15 in CostCalcs sheet; M126, N126 & O126 in TLS prophylaxis sheet
Changed OS HR 0.555 (mean OS HR, CS adjusted anchored MAIC)	[redacted]	[redacted]	[redacted]	-£149,447	1.475	Dominated	C142
Changed OS HR to 0.201 (lower 95% CI, CS adjusted anchored MAIC)	[redacted]	[redacted]	[redacted]	-£84,647	3.269	Dominated	C142
Changed OS HR to 1.534 (upper 95% CI, adjusted anchored MAIC)	[redacted]	[redacted]	[redacted]	-£172,056	-0.88	£195,564	C142
Changed OS HR 1.075 (OS HR, ERGs IC)	[redacted]	[redacted]	[redacted]	-£163,766	-0.027	£6,117,189	C142
Changed OS HR to 0.423 (lower 95% CI, ERGs IC)	[redacted]	[redacted]	[redacted]	-£144,557	.999	Dominated	C142

Changed OS HR to 2.728 (upper 95% CI, ERGs IC)	[redacted]	[redacted]	[redacted]	-£183,238	-2.025	£90,504	C142
Changed PFS HR to 1.429 (PFS HR, ERGs IC)	[redacted]	[redacted]	[redacted]	-£135,650	2.599	Dominated	C132
Changed PFS HR to 0.780 (PFS HR, lower 95% CI, ERGs IC)	[redacted]	[redacted]	[redacted]	-£135,650	2.599	Dominated	C132
Changed PFS HR to 2.615 (PFS HR, upper 95% CI, ERGs IC)	[redacted]	[redacted]	[redacted]	-£89,952	2.599	Dominated	C132
Changed PFS and OS to joint Gamma	[redacted]	[redacted]	[redacted]	-£142,716	2.884	Dominated	C112 & C113
Changed PFS and OS to joint-Gamma, used mean HR for PFS and OS from ERGs IC	[redacted]	[redacted]	[redacted]	-£322,979	-0.39	£827,252	C112 & C113; C142 & C132
Changed PFS and OS joint-Gamma, lower 95% CI for PFS from ERGs IC	[redacted]	[redacted]	[redacted]	-£142,716	2.884	Dominated	C112 & C113; C132
Changed PFS and OS fits to joint-Gamma, upper 95% CI for PFS HR from ERGs IC	[redacted]	[redacted]	[redacted]	-£142,716	2.884	Dominated	C112 & C113; C132
Changed PFS and OS joint-Gamma, lower 95% CI of OS from ERGs IC	[redacted]	[redacted]	[redacted]	-£162,911	2.197	Dominated	C112 & C113; C142
Changed PFS and OS to joint-Gamma, upper 95% CI of OS HR from ERGs IC	[redacted]	[redacted]	[redacted]	-£201,819	-1.846	£109,308	C112 & C113; C142
Changed PFS and OS to joint-Log-logistic	[redacted]	[redacted]	[redacted]	-£137,588	3.066	Dominated	C112 & C113
Changed PFS and OS HR to 1.429 and 1.075 on joint Weibull (ERGs IC)	[redacted]	[redacted]	[redacted]	-£279,766	-0.354	£790,988	C132 & C142
Applied ERGs preferred base-case model to del(17p)/TP53 population	[redacted]	[redacted]	[redacted]	-£269,728	-0.376	£718,043	B2; C112 & C113, C132 & C142
Applied ERGs preferred base-case model to non-del(17p)/TP53 population	[redacted]	[redacted]	[redacted]	-£343,718	-0.393	£873,858	B2; C112 & C113, C132 & C142

* unless stated; IC = indirect comparison

Table 52: Further exploratory analyses undertaken by the ERG for the comparison with IDELA+R

ERG exploration	List price comparisons			Net price comparisons			Cell changes in GEN SETTINGS sheet in economic model*
	Incremental costs	Incremental QALYs	ICER	Incremental costs	Incremental QALYs	ICER	
CS corrected base-case	[redacted]	[redacted]	[redacted]	£11,726	3.358	£3,492	-
Changed mean age, % male and % del(17p) to Study 116 figures	[redacted]	[redacted]	[redacted]	£13,815	3.084	£4,480	C8, C9 & C11
Changed routine care costs to figures in TA487	[redacted]	[redacted]	[redacted]	£18,468	3.358	£5,499	G15 & H15 in CostCalcs sheet
Changed TLS prophylaxis costs to estimates used in TA487	[redacted]	[redacted]	[redacted]	£11,958	3.358	£3,561	M126, N126 & O126 in TLS prophylaxis sheet
Changed all AE costs: (Anaemia, Anaemia (Autoimmune haemolytic) , Neutropenia, Pneumonia, Thrombocytopenia) to estimates in NICE TA439	[redacted]	[redacted]	[redacted]	£12,150	3.358	£3,618	C34 to C39
Changed TLS prophylaxis costs and routine care costs to figures in TA487; and AE costs to the figures in TA439	[redacted]	[redacted]	[redacted]	£19,123	3.358	£5,694	C34 to C39; G15 & H15 in CostCalcs sheet; M126, N126 & O126 in TLS prophylaxis sheet
Changed PFS to joint-Gamma	[redacted]	[redacted]	[redacted]	£8,100	3.431	£2,361	C112
Changed OS to joint-Gamma	[redacted]	[redacted]	[redacted]	£14,337	3.617	£3,963	C113
Changed PFS and OS to joint-Gamma	[redacted]	[redacted]	[redacted]	£10,711	3.69	£2,903	C112 & C113
Changed PFS and OS to joint-Log-logistic	[redacted]	[redacted]	[redacted]	£7,737	3.837	£2,017	C112 & C113
Changed OS hazard ratio to 1 (equal efficacy assumption between VEN+R and IDELA+R	[redacted]	[redacted]	[redacted]	-£14,944	0.512	Dominated	C143

* unless stated