Venetoclax with obinutuzumab for untreated chronic lymphocytic leukaemia ID1402

Produced by: Warwick Evidence

Authors
Daniel Gallacher, Research Fellow, Warwick Evidence
Lazaros Andronis, Associate Professor, Warwick Evidence
Jill Colquitt, Senior Researcher, Effective Evidence LLP
Mandana Zanganeh, Research Fellow, Warwick Evidence
Emma Loveman, Senior Researcher, Effective Evidence LLP
Samantha Johnson, Information Specialist, Warwick Medical School
Scott Marshall, Consultant Haematologist, South Tyneside and Sunderland NHS Foundation Trust
Daniel Todkill, Clinical Research Fellow, Warwick Medical School
Anna Brown, Information Specialist, Warwick Evidence
Hema Mistry, Associate 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 15th April 2020

Source of funding: This report was commissioned by the NIHR Systematic Reviews Programme as project number 129899.

Declared competing interests of the authors

Dr Scott Marshall received sponsorship from AbbVie to attend educational conferences, and he has been paid to provide educational talks for clinicians and to attend advisory boards. He has no financial links or contracts with AbbVie. He has received similar payments/ sponsorship from most other pharmaceutical companies including Roche,

Astrazeneca, BMS, Pfizer, Novartis, Celgene, Merck, Sanofo, Johnson & Johnson and Takeda.

Acknowledgements

Professor Aileen Clarke 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 with obinutuzumab for untreated chronic lymphocytic leukaemia. A Single Technology Appraisal. Warwick Evidence, 2020.

Contributions of authors

Daniel Gallacher (Research Fellow) reviewed and critiqued the statistical analysis and undertook additional analyses; Lazaros Andronis (Associate Professor) and Mandana Zanganeh (Research Fellow) reviewed and critiqued the cost-effectiveness evidence and undertook additional analyses; Jill Colquitt (Senior Researcher) and Emma Loveman (Senior Researcher) reviewed and critiqued the clinical effectiveness evidence; Samantha Johnson (Information Specialist) critiqued the company searches and undertook additional searches; Anna Brown (Information Specialist) provided additional information specialist support; Scott Marshall (Consultant Haematologist) provided expert clinical advice; Daniel Todkill (Clinical Research Fellow) provided clinical advice; and Hema Mistry (Associate Professor) co-ordinated the project and the report, and reviewed the draft and final reports.

Word count: 44,065 (including references)

Contents

List of Tables

List of Figures

Glossary

AE	adverse event
AIC	Akaike information criteria
B-CLL	B-Cell Chronic Lymphocytic Leukemia
BIC	Bayesian information criteria
BNF	British National Formulary
BR	bendamustine and rituximab
BSH	British Society of Haematology
CDSR	Cochrane Database of Systematic Reviews
CEM	cost-effectiveness model
CHMP	Committee for Medicinal Products for Human Use
CI	confidence interval
CIRS	cumulative illness rating scale
Clb	chlorambucil
CLL	chronic lymphocytic leukaemia
COPD	chronic obstructive pulmonary disease
CrCI	creatinine clearance
CR	complete response
CRi	complete response with incomplete bone marrow recovery
CS	company submission
CSR	clinical study report
del(17p)	deletion of the short arm of chromosome 17
DOR	duration of response
EFS	event-free survival
EMA	European Medicines Agency
EORTC QLQ-C30	European Organization for Research and Treatment of Cancer Quality of Life Questionnaire Core 30
EOT	end of treatment assessment
EQ-5D-3L	EuroQol 5 dimensions
ERG	Evidence Review Group
ECOG	Eastern Cooperative Oncology Group
FCR	fludarabine, cyclophosphamide and rituximab
FTD	fixed treatment duration
GClb	chlorambucil and obinutuzumab
HR	hazard ratio
HRQoL	health-related quality of life
HSUV	health state utility value
ICER	incremental cost-effectiveness ratio
IGHV	immunoglobulin heavy chain gene variable region
IRC	independent review committee
ITT	intention-to-treat
LYs	life years
KM	Kaplan-Meier
MAIC	matching adjusted indirect comparison
MDASI-CLL	M.D. Anderson Symptom Inventory-CLL
MRD	minimal residual disease
NE	not evaluable
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
NMB	net monetary benefit
NR	not reported
ORR	overall response rate

OS	overall survival
PAS	patient access scheme
PD	progressive disease
PFS	progression free survival
PPS	post-progression survival
PR	partial response
PROs	patient reported outcome measures
PS	performance status
PSA	probabilistic sensitivity analysis
QALYs	quality-adjusted life years
QOL	quality of life
RCT	randomised controlled trial
SAE	serious adverse event
SLL	small lymphocytic lymphoma
SLR	systematic literature review
SPC	summary of product characteristics
TA	technology appraisals
TEAE	treatment-emergent adverse event
TOT	time on treatment
TONT	time on next treatment
TTE	time-to-event
TTNT	time to the next anti-leukemic treatment
TLS	tumour lysis syndrome
UK	United Kingdom
VenG	venetoclax with obinutuzumab
WTP	willingness-to-pay

1 SUMMARY

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

The National Institute for Health and Care Excellence (NICE) scope for this appraisal is the clinical and cost-effectiveness of venetoclax with obinutuzumab (VenG) within its marketing authorisation for untreated chronic lymphocytic leukaemia (CLL).

The population in the company's submission (CS) is people with untreated CLL with coexisting conditions that make fludarabine and bendamustine (FCR/BR) based therapy unsuitable. This is the population in the pivotal CLL14 trial and reflects the company's anticipated positioning of VenG in the National Health Service (NHS) treatment pathway. However, this population is narrower than the anticipated marketing authorisation and may not be wholly generalisable to the population that UK clinicians wish to use VenG for.

The CS considers two key subgroups: those without a del(17p) or TP53 mutation and those with a del(17p) or TP53 mutation. The algorithm for identifying the del(17p)/TP53 mutation subgroup differs between the clinical and cost-effectiveness sections, resulting in differing sample sizes. The rationale for using different algorithms is not adequately justified by the company, although it has little impact on the results.

The comparators in the NICE scope are considered for the two subgroups. The submission includes obinutuzumab plus chlorambucil (GClb) for those without a del(17p) or TP53 mutation and ibrutinib for those with a del(17p) or TP53 mutation. Other scoped comparators are excluded with justification, which the Evidence Review Group (ERG) agrees with.

1.2 Summary of clinical effectiveness evidence submitted by the company

The company provided data to the ERG in the following four submissions:

- The original CS and clinical study report (CSR): data-cut August 2018 (28.1 months median follow-up)
- The clarification responses and CSR Corrigendum: data-cut August 2018, correcting errors in the original CS Figure 15 for subgroup analysis of del(17p)/TP53 mutation
- The CS addendum and CSR supplement: data cut August 2019 (39.6 months median follow-up)
- The CS addendum clarification responses: data-cut August 2019

The CS presents evidence from one multi-centre randomised controlled trial (RCT) investigating the effectiveness and safety of VenG in people with previously untreated CLL with co-existing medical conditions. The comparator in the CLL14 trial was chlorambucil and obinutuzumab (GClb). Presence of coexisting conditions was defined by a total cumulative illness rating scale (CIRS) of >6 or creatinine clearance <70ml/min. The trial included 368 participants without del(17p)/TP53 mutation, 49 with del(17p)/TP53 mutation and 15 with missing data (clarification A4). Randomisation led to 216 participants in each treatment arm.

Follow-up of the CLL14 trial is ongoing and the CS addendum presents results from a data cut after a median of 39.6 months; when all participants had completed 12 cycles of treatment (August 2019). At this point, 177 participants remained in follow-up in the VenG arm and 178 in the GClb arm. The key outcomes are summarised below. Median progression free survival (PFS) or overall survival (OS) had not been reached in the VenG arm at the time of the analysis.

- Investigator assessed PFS (trial primary outcome) demonstrated superiority of VenG with a hazard ratio of 0.31 (95% confidence interval (CI) 0.22 to 0.44, p<0.0001). [redacted] deaths had occurred in each arm at the time of the latest follow-up. The hazard ratio for OS was 1.03 (95% CI 0.60 to 1.75, p=0.92), suggesting no difference between VenG and GClb, although there is a degree of uncertainty around the estimate.
- Response outcomes were assessed at end of treatment (3 months after a patient received their last treatment dose). A formal analysis of complete response (CR) and CR with incomplete bone marrow recovery (CRi) combined, demonstrated a difference in response rate of 26.4% in favour of VenG (95% CI 17.4% to 35.4%, p<0.0001).
- The stratified hazard ratio of duration of response (DOR), defined as the time from the first occurrence of a response until disease progression or death, was [redacted] (95% CI [redacted]) and separation of the Kaplan-Meier curves indicated a superior DOR in favour of VenG.
- Time to the next anti-leukemic treatment (TTNT) defined as the time between the date of randomisation and the date of a patient receiving a second line therapy or death also suggested that VenG has a significantly lower hazard rate of next treatment or death than GClb (stratified hazard ratio 0.51 (95% CI 0.34 to 0.78, p = 0.0012)). See discussion below regarding a potential interpretation issue with these data.

- The CS presents minimal residual disease (MRD) as a secondary trial outcome, although this was not a NICE scoped outcome. MRD was measured in both blood and bone marrow at various times during the trial. The main secondary outcome was rate of MRD negativity in blood at the 3 month post-treatment follow-up. The company reports that VenG achieved 75.5% MRD negativity compared to 35.2% in the GClb arm 3 months after treatment completion, but this reduced to 47.2% and 7.4%, respectively, 18 months after treatment completion or early termination. MRD negativity in bone marrow at 3 months post-treatment showed lower rates of negativity for both arms than the blood measurements (VenG 56.9%; GClb 17.1%), this was not measured 18 months after treatment completion.
- Health-related quality of life (HRQoL) was assessed with the EuroQol 5 dimensions [EQ-5D-3L]; European Organization for Research and Treatment of Cancer Quality of Life Questionnaire Core 30 [EORTC QLQ-C30] and symptoms were assessed with the M.D. Anderson Symptom Inventory-CLL [MDASI-CLL]. All showed a [redacted] from baseline score [redacted], but there were [redacted] between the arms of CLL14 at any point in the follow-up.
- The majority of participants experienced at least one treatment-emergent adverse event (TEAE); 14.6% and 15.9% of the VenG and GClb groups, respectively, discontinued a treatment for TEAEs. The most common Grade 3-4 adverse event was neutropenia, occurring in [redacted] (VenG) and [redacted] (GClb), respectively. Other common Grade 3-4 adverse events included thrombocytopenia, infusion related reaction and febrile neutropenia.
- In total [redacted]% of VenG participants and [redacted]% of GClb participants experienced at least one serious adverse event (SAE). The most frequently reported SAEs were febrile neutropenia, pneumonia, infusion-related reaction and pyrexia. In total [redacted]% (VenG) and [redacted]% (GClb) of participants had an adverse event that resulted in death.
- Tumour Lysis Syndrome (TLS) was reported in three VenG treated participants and in five GClb treated participants.

A naïve indirect comparison was made between CLL14 and three separate studies to compare VenG with ibrutinib for people with previously untreated CLL and del(17p)/TP53 mutation. The company's preferred study for this analysis was a retrospective cohort study of people with CLL treated with ibrutinib (Mato 2018). The study had a subgroup with del(17p) (n=110). Another study (Ahn 2018) was a single arm study of ibrutinib for CLL which reported a subgroup who were untreated and who had del(17p)/TP53 mutation (n=35); this indirect

comparison was undertaken as a sensitivity analysis in the CS. A comparison with ALLIANCE data was provided in clarification A23 and updated in the CS addendum.

- For the main comparison (Mato 2018), fitting a Cox proportional hazard model to the data produced a PFS hazard ratio of 0.660 (95% CI 0.270 to 1.615, p = 0.363). The confidence intervals are wide illustrating how uncertain the results are. Fitting another Cox model to the OS data produced a hazard ratio of 0.841 (95% CI 0.301 to 2.352, p=0.741). Again, the confidence intervals are wide.

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

The whole trial population of CLL14 is used as evidence for the subpopulation of people without del(17p)/TP53 mutation compared with GClb. This evidence includes a proportion of participants with del(17p)/TP53 mutation. The subgroup with del(17p)/TP53 mutation is compared via a naïve indirect comparison with ibrutinib monotherapy. The ERG notes therefore there is some double counting of participants.

The ERG generally agrees with the CS assessment that the trial has a low risk of bias for most domains, however we note the performance bias and detection bias inherent in an open label trial. We also consider that there is a risk of bias due to selective outcome reporting. In addition, although the arms are balanced with respect to key prognostic factors, there are some differences between groups in baseline comorbidities.

The details of, and reasons for, dose modifications of venetoclax, chlorambucil and obinutuzumab were not consistently reported in the CS.

The CS includes a non-scoped outcome, MRD rate. The ERG clinical advisor confirms that undetectable MRD is an important surrogate endpoint, particularly in bone marrow, and that there is a relationship between undetectable MRD and final outcomes in CLL, although recent evidence suggests the relationship between MRD and outcomes following venetoclax needs further validation.

The trial follow-up is ongoing and the data presented are from a data-cut that was not the originally planned primary analysis point. Data are therefore immature and it is not possible to draw conclusions for all of the specified outcomes.

The company does not present an analysis of whether the hazard ratio, which assumes proportionality of the hazard rates between the two trial arms, is a suitable outcome when reporting results of PFS. In the cost-effectiveness section, the company concludes that proportionality is not held and this would suggest that the estimate of the hazard ratio is not an accurate representation of the benefit of VenG on PFS. The ERG also notes that the analyses of DOR and event-free survival were performed without an assessment of proportional hazards. The analysis for OS was presented without a discussion of proportional hazards in the clinical effectiveness section but the assumption was investigated in the cost-effectiveness section of the CS.

It is unclear whether the inclusion of OS events introduces bias into the analysis of TTNT, as the company treats deaths as events in the TTNT analysis rather than as censored observations as stated in the original CLL14 trial protocol (though is consistent with later versions). The ERG therefore suggests caution in the interpretation of the TTNT results. Analysis of the TTNT outcome where death events were censored, provided in response to a request by the ERG, produced a hazard ratio of [redacted], but without confidence intervals or test of statistical significance (addendum clarification response).

Generalisability may be limited due to the restricted population reported in the evidence including only those with comorbidities in whom FCR/BR based therapy is unsuitable rather than the NICE scoped population. The trial was international and undertaken in 21 countries, there were 6 United Kingdom (UK) sites and 8 UK participants; there may be reduced generalisability to the NHS population because of this. Some patients in the trial may have been eligible for FCR/BR therapy in the UK and therefore may be slightly healthier than the UK population unsuitable for FCR/BR. Also, in practice the decision about whether a patient is suitable for FCR/BR therapy is based on an end-of-bed assessment, and is not necessarily well-reflected in the CIRS cut off used in CLL14.

The CS undertook a feasibility assessment to determine the suitability of available data for an indirect comparison to ibrutinib in the del(17p)/TP53 mutation subgroup. The ERG agrees that an anchored comparison is not possible and that an unanchored match adjusted indirect comparison is also not ideal.

Three studies with subgroups of relevance were identified and compared with CLL14 via a naïve comparison. The ERG considers it uncertain whether the subgroups in these studies are comparable with the CLL14 del(17p)/TP53 mutation subgroup. In addition, the ERG notes that there are a number of inaccuracies in the description of the Ahn 2018 study by the CS. There is heterogeneity between these studies and CLL14 in terms of study design, eligibility criteria, outcomes and possible heterogeneity in baseline characteristics.

The company performed the indirect comparison using hazard ratios but the data are currently insufficient to conclude whether hazard ratios accurately capture the differences between the treatments. Also, there was no patient level data for the comparator participants and the data were obtained from digitising graphs which is another source of uncertainty.

Overall, caution is recommended in the interpretation of these indirect comparisons.

1.4 Summary of cost effectiveness submitted evidence by the company

Evidence on cost-effectiveness was received on the following occasions:

- The original CS, which was based on CLL14 evidence from the August 2018 data cut.
- Responses to ERG's request for clarifications on the original CS.
- The CS addendum, presenting an updated analysis taking into account a new data cut-off (August 2019)
- Responses to ERG's request for clarifications on the CS addendum.

Economic models were made available with the original submission (based on the August 2018 data-cut), with the CS addendum and with the company's response to clarification queries on the original CS and CS addendum.

The submitted evidence pertains to two distinct subgroups: 1) patients without a del(17p)/TP53 mutation and 2) those patients with a del(17p)/TP53 mutation. Treatments compared for the first subgroup are VenG and GClb. Treatments compared for the second subgroup are VenG and ibrutinib.

The company carried out a systematic literature review of cost-effectiveness evidence, aiming to identify studies on previously untreated CLL that reported relevant economic evaluations,

health-related quality of life (HRQoL) or costs and use of health care resources. The review identified 43 relevant economic evaluations, 20 studies providing information on HRQoL and 16 studies giving estimates of healthcare resource use and costs. The company concluded that none of the identified economic evaluations pertained to the exact decision problem of interest in this submission. Information identified through the review, including evidence from completed NICE technology appraisals in CLL, was used in the economic analysis.

The economic model submitted follows a partitioned survival approach and comprises three health states: (i) progression-free survival, (ii) post-progression survival and (iii) death. The distribution of the patient population within each of the three health states at each point in time is guided by extrapolated progression-free survival (PFS) and overall survival (OS) curves. Time-to-next treatment curves were also used to calculate the point in time when subsequent treatment was initiated. Death due to causes other than CLL (i.e. background mortality) was guided by age-adjusted and sex-adjusted mortality risk values drawn from UK life tables. The model adopted a NHS and personal social services perspective, uses a 28-day cycle length, has a time horizon of 30 years and discounted future costs and benefits at 3.5% per annum (in the base-case analysis).

The same model structure was used to evaluate the cost-effectiveness of the compared treatments (VenG vs GClb in patients without del(17p)/TP53 mutation; VenG vs ibrutinib in patients with del(17p)/TP53 mutation). The algorithm used to categorise patients to groups according to mutation status differed between the clinical study report analysis and the cost-effectiveness modelling analysis.

Time to event parameter estimates used in the model were obtained from the CLL14 trial (August 2018 data cut in the original CS, August 2019 data cut in subsequent addendum). For the non-del(17p)/TP53 mutation population (VenG vs GClb), PFS and OS were informed by data from the CLL14 trial and were parameterised using an independent model (log-logistic) and a dependent model (exponential), respectively. For OS, the company used the predicted curve for the GClb arm to represent OS for both arms. Time-to-next treatment (TTNT) was extrapolated using an independent (log-logistic) model applied to CLL14 data for both VenG and GClb arms. All curves were constrained such that their hazard rates could not fall below background mortality. For the del(17p)/TP53 mutation population (VenG vs ibrutinib), the company pointed out that the limited evidence of ibrutinib in the untreated CLL with del(17p)/TP53 mutation population made network meta-analyses and matched adjusted indirect comparison

unfeasible. A naïve comparison of VenG versus ibrutinib was performed using a published study by Mato et al.

Preference-based quality of life (utility) values for different health states were collected in the CLL14 trial using the EuroQol EQ-5D-3L instrument. However, the company considered that the utility estimates were notably higher than those accepted in previous appraisals and published UK age-adjusted general population values. Instead, a decision was made to use health state utilities values for the pre-progression (PFS) and post-progression (PPS) states from the available literature. Utility values relating to pre-progression status were further broken down by treatment receiving status (on treatment, off treatment) and type of treatment (intravenous or oral).

The following key categories of resource use and costs were included in the company's analysis: (i) intervention and comparators' costs (including treatment acquisition and administration costs, routine care costs, tumour lysis syndrome (TLS) monitoring costs and subsequent treatment costs), (ii) costs related to adverse events, and (iii) terminal care costs. Unit costs of drugs comprising VenG and its comparators were sourced from the British National Formulary (BNF). Administration costs were included in the model for the treatments delivered intravenously. Routine care and monitoring costs included services such as scans, blood tests, transfusions and consultations. The cost of TLS prophylaxis was calculated based on an algorithm that categorised patients by risk of developing TLS according to data observed in the treated CLL14. Subsequent treatment costs were calculated according to the type of subsequent treatment mix received, the point in time when subsequent treatment would be initiated and the length of time over which subsequent treatment would be administered. In the non-del(17p)/TP53 mutation population, input for these calculations was derived from time-to-event data observed in CLL14 (TTNT and OS). In the del(17p)/TP53 mutation population, the proportion of ibrutinib patients who receive subsequent treatment was calculated as the difference in the ibrutinib PPS duration and OS curves.

On the basis of list prices for all treatments, the company reported the following results. In the non-del(17p)/TP53 mutation population, VenG is associated with a greater number of QALYs and lower costs than GClb, thus, VenG is dominant versus GClb. In this population, VenG resulted in a positive net monetary benefit (NMB) of [redacted] at a willingness-to-pay threshold of £30,000 per QALY. In the del(17p)/TP53 mutation population, VenG is associated with a lower number of QALYs and lower costs versus ibrutinib. In this population, VenG was associated with

a positive NMB of [redacted] at a willingness-to-pay threshold of £30,000 per QALY. Sensitivity and scenario analyses reported by the company showed that, on the whole, results are robust to alternative values and assumptions.

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

The following key points in relation to cost-effectiveness evidence presented by the company have been discussed in the ERG's critique and are summarised here:

- Systematic literature reviews carried out by the company to identify existing evidence on economic evaluations, costs and HRQoL were comprehensive. The ERG accepts that no directly relevant economic evaluations are available and agrees that developing a *de novo* economic model tailored to the requirements of the specific final scope and decision problem was necessary.
- The ERG believe that the type and structure of the submitted model (three state partitioned survival model) is appropriate for the purposes of the condition investigated and adequate for the decision problem considered in this appraisal. The pathway employed in the model is, in general, in line with expectations around the clinical progression of the disease, while the structure of the model is generally suitable for capturing and quantifying key costs and health outcomes associated with the compared treatments.
- More broadly, the analytic methods used in the economic analysis (evaluated time horizon, discounting, evaluation of costs and outcomes) are generally in line with the NICE Guide to Methods of Technology Appraisal and previous NICE TAs.
- Different approaches for categorising CLL14 patients according to mutation status were used in the clinical study report analysis and the economic model, resulting in differences in the numbers of patients included in the defined populations with and without del(17p)/TP53 mutation. However, the data resulting from the different categorisations had a small, inconsequential impact on the final cost-effectiveness results.
- Immaturity of data (in VenG vs. GClb), reliance on an unadjusted naïve indirect comparison (for VenG vs ibrutinib) and the uncertainty arising as a result have an inevitable effect on cost-effectiveness calculations. Time-to-event data are drivers of incremental costs and outcomes in the decision model. Limitations in currently available data make it difficult to draw a complete and reliable picture of each treatment's

effectiveness and they inevitably affect the final cost-effectiveness results. The ERG has identified extrapolations that, we believe, are more plausible and appropriate; these have been incorporated in the ERG's preferred base-case analysis.

- Health state utility values were sourced from the literature, rather than from the EQ-5D data collected in the CLL14 trial. The justification for not using CLL14 trial observations is considered to be reasonable. QALY decrements due to adverse events were appropriately applied. In response to clarification questions, the company offered a more pragmatic reflection of utility values in the pre-progression health state, which takes into account whether patients are off or on treatment and the type of treatment received (intravenous or oral). However, the ERG consider the utility value assigned to reflect the 'progression-free, off treatment' status to be problematic. An alternative value has been put forward as a more plausible estimate in the ERG's base-case analysis.
- A number of resource use components and their relevant costs were identified and taken into account in the cost calculations. These included acquisition and administration costs for first and second line treatments, routine care and tests, cost of TLS prophylaxis and terminal care costs. Elements of the calculations and methods used are in line with previous NICE Technology Appraisals in CLL. An inconsistency in the calculation of subsequent treatment costs for patients on ibrutinib first line treatment has been pointed out.
- The company took a number of steps to validate the submitted economic model. Additional checks were carried out by the ERG. The ERG agree that steps undertaken by the company to ensure the validity of the model are appropriate. Putting aside limitations in the analysis due to data immaturity and unavailability, the ERG deem the model's validity to be, on the whole, sound.
- The company carried out probabilistic, deterministic and scenario analyses. Issues identified around the specified level of uncertainty across a range of diverse parameters were raised in the ERG's clarification questions and corrections were made by the company. In general, sensitivity analyses suggested that the results are robust to a wide range of alternative values and approaches.

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

1.5.1 Strengths

Evidence presented by the company presented the following strengths:

- The review methods employed in the company systematic literature review were appropriate and there is a low risk of systematic error in the results of the review.
- The included trial (CLL14) was well designed and has a low risk of bias within the limits of an open label design. It is suitably powered to answer the primary hypothesis.
- The model type and structure were appropriate for the decision problem.
- Where available, key evidence on treatment effectiveness was drawn from the CLL14 trial.
- Resource use and costs calculations were in agreement with NICE technology appraisals in CLL.
- Extensive sensitivity and scenario were carried out to assess the robustness of the results to different assumptions, methods and parameter values.

1.5.2 Weaknesses and areas of uncertainty

Evidence submitted by the company presented the following key weaknesses:

- The CS used the whole trial population from the CLL14 trial, which included those with and those without del(17p)/TP53 mutation, for the subpopulation of people without the del(17p)/TP53 mutation comparing VenG with GClb. It is unclear what effect this may have on the results, although the ERG notes that the numbers with del(17p)/TP53 mutation were small. There is also double counting of participants with del(17p)/TP53 mutation as these are then used separately in the analysis comparing with ibrutinib.
- The CLL14 trial is ongoing and data are immature for some of the key outcomes of relevance to the decision problem. The assessment of proportional hazards is not clearly reported for some outcomes. The reasons for, and level of, dose reductions or alterations of the treatments within the CLL14 trial are not reported for all treatments consistently and the impact of these modifications is uncertain.
- There is no head-to-head comparison between VenG and ibrutinib and a naïve indirect comparison was undertaken. The results of this indirect comparison are very uncertain owing to the methodological approaches used and likely heterogeneity between populations; this is reflected in the wide confidence intervals seen and results should therefore be interpreted with caution.
- Immaturity of data (in VenG vs. GClb), use of a naïve indirect comparison (for VenG vs ibrutinib) and the uncertainty arising as a result, mean that key time-to-event data which are drivers of incremental costs and outcomes in the decision model may not be

appropriate. To the extent possible, the ERG has identified extrapolations that are deemed to be more appropriate.

- Due to unexpectedly high preference-based health related quality of life (EQ-5D-3L) values observed in the CLL14 trial, health state utilities were sourced from the literature. The ERG disagreed with the value chosen to reflect utility in patients who had not progressed and were off treatment, which is deemed to be inappropriately high.

1.6 Summary of exploratory and sensitivity analyses undertaken by the ERG

To address issues identified in the economic analysis submitted by the company, the ERG implemented changes that formed the ERG's preferred base-case analysis. Amendments were made to the utility value for the 'pre-progression, off treatment' status and in time-to-event parameters and extrapolations in both the non-del(17p)/TP53 and del(17p)/TP53 mutation populations.

In the non-del(17p)/TP53 mutation population, implementing the ERG's preferred base-case resulted in reductions in incremental costs and QALYs compared to the company's base-case values, leading to an overall reduction in NMB (at list prices for all treatments and a willingness-to-pay value £30,000 per QALY) by [redacted] compared to the company's base-case results ([redacted] vs [redacted]). In both the company's base-case analysis and the ERG's base-case analysis, VenG was dominant against GClb.

In the del(17p)/TP53 mutation population, implementing the ERG's preferred base-case resulted in reductions in both incremental costs (cost savings) and QALYs (VenG vs ibrutinib) compared to the company's base-case values, leading to an overall reduction in NMB (at list prices for all treatments and a willingness-to-pay value £30,000 per QALY) by approximately [redacted] compared to the company's base-case results ([redacted] vs [redacted]). The ICER for this comparison falls within the south-west quadrant of the cost-effectiveness plane reflecting cost savings per QALY forgone.

Additional scenario analyses carried out by the ERG led to results that agreed in direction with the results of the company's and the ERG's base-case analyses. No additional, non-quantifiable variables that may have a consequential change in the results were identified.

2 BACKGROUND

2.1 Critique of company's description of underlying health problem

The CS provides an overview and description of the epidemiology of chronic lymphocytic leukaemia (CLL) (CS B.1.3.1), noting that it is a clonal disease [i.e. involving development of identical cells] of unknown aetiology characterised by the accumulation of mature B cells in blood, lymph nodes, spleen liver and bone marrow. This leads to leucocytosis [increase in white blood cells], lymphadenopathy [abnormal lymph nodes], hepatosplenomegaly [swelling of liver and spleen], anaemia [decrease in red blood cells or haemoglobin], thrombocytopenia [platelets deficiency], neutropenia [reduced neutrophils leading to increased susceptibility to infection], bone marrow failure [insufficient red blood cells, white blood cells or platelets produced] and symptoms as described below.

The CS describes some of the genetic abnormalities that can be identified in CLL, namely mutation of the TP53 gene via deletion of the short arm of chromosome 17 [del(17p)] (which contains the TP53 gene) or mutation of the TP53 gene sequence. About 5-10% of CLL patients have TP53 dysregulation at diagnosis[1] and it is associated with poor prognosis; chemoimmunotherapy is ineffective in these patients.[2]

The CS does not describe immunoglobulin heavy chain gene variable region (IGHV) mutation. This is a known prognostic marker, with IGHV-mutated CLL associated with better prognosis and slower growing disease. Retrospective studies suggest that patients with mutated IGHV can experience prolonged remissions with chemotherapy.[2] The ERG clinical advisor states that it is now increasingly tested for in routine clinical care within the NHS.

The CS notes that CLL is the most common of the chronic leukaemias. It reports the European age-standardised incidence in the UK as 6.0 per 100,000 in 2016, with 3,412 new cases in England and Wales alone.[3] It is more common in men than women, with a ratio of 1.7: 1. The risk increases with age; with 42% of new cases in people aged 75 and over and the highest incidence in women aged 85 to 89 and men aged over 90.[3]

The CS states that most CLL patients are older than 70 and have relevant coexisting conditions, but does not describe details of frequency and type of coexisting conditions in CLL. Co-morbidities can affect an individual's fitness for chemotherapy, and treatment options differ for those fit for chemotherapy (or 'suitable for') and those unfit (unsuitable) for chemotherapy (see

below). However, as the CS acknowledges, there is no optimal strategy or agreed co-morbidity assessment tool to determine fitness for chemotherapy.[2] According to the ERG clinical expert, in UK practice clinicians rely on 'end of the bed assessment' which is difficult to quantify. In addition to specific co-morbidities, other factors such as poor performance status / exercise capacity, poor bone marrow reserve, contraindications to treatment and desire to avoid intravenous chemotherapy or regular hospital attendance for treatment may also make a patient unsuitable for certain treatments.

The CS describes the disease burden (CS B.1.3.2). Most patients are asymptomatic at diagnosis and are diagnosed by chance through routine blood tests. Symptoms that can appear as the disease progresses include swollen lymph nodes, recurrent infections and systemic symptoms (fatigue, loss of appetite, weight loss, night sweats and shortness of breath when exercising).

The CS describes the impact of CLL on the patient's quality of life and ability to work. The CS reports findings from a large prospective survey of people with CLL which showed that disease progression has a negative impact on health-related quality of life (HRQoL).[4] The study also showed that people with CLL have lower emotional wellbeing than the general population and people with other cancer types, although this comparison was made with historical controls.[4] However, the CS does not report that the survey also found similar physical, social/family, functional, and overall quality of life (QOL) scores for CLL patients and those from published population norms.[4]

The CS cites evidence that an additional burden for people with CLL is the impact on the ability to work.[5] This evidence source is a guide for patients produced by the Leukaemia Care charity and discusses how people with CLL may require temporary sick leave, reduction in working hours or need reasonable adjustments to be made at work. The CS hypothesises that these factors may have an impact on finances and emotional burden of people with CLL, which seems reasonable, but the CS does not present any evidence to support this.

Overall, the ERG considers that the company's description of the underlying health problem is appropriate and relevant to the decision problem under consideration.

2.1 Critique of company's overview of current service provision

The CS describes the current UK CLL clinical pathway of care (CS B.1.3.4 and summarised in CS Figure 1), including diagnosis and staging, initiation of treatment and determining fitness status for chemotherapy. Treatments recommended by NICE and by British Society of Haematology (BSH) guidelines[2] are outlined in CS Table 2 together with their relevance to the current submission. Treatment of previously untreated CLL patients is described for the following groups:

- *Fit patients without del(17p)/TP53 mutation:*

The CS refers to BSH guidelines[2] for treatment with fludarabine, cyclophosphamide and rituximab (FCR) and bendamustine and rituximab (BR), which are listed as treatment options for this group in the NICE scope. However, the NICE scope also lists treatment with chlorambucil with or without rituximab, which is not mentioned by the CS. The company considers that this population is not relevant to the submission, as the population of the pivotal CLL14 trial had characteristics that would typically make then unsuitable for FCR and BR. However, the ERG clinical expert considers that this patient group would be relevant to this appraisal (see section 3.1).

- *FCR/BR unsuitable patients without del(17p)/TP53 mutation:*

In line with BSH guidelines[2] and the NICE scope, the CS notes that obinutuzumab with chlorambucil (GClb) is the standard of care in this group.

- *Patients with del(17p)/TP53 mutation:*

The CS notes that NICE recommends ibrutinib monotherapy for people for whom chemoimmunotherapy is unsuitable, and idelalisib with rituximab for people who cannot have other therapies. The CS argues that the latter has now been superseded by ibrutinib due to a higher risk of infection and death than other therapies. The ERG clinical expert agrees with this statement. The CS notes that the BSH guidelines[2] also recommend ibrutinib in this population.

The ERG considers that the company's overview of current service provision is appropriate and relevant to the decision problem under consideration.

Unmet need

The company describes the unmet need for treatment options for previously untreated CLL (B.1.3.5), particularly for those with del(17p)/TP53 mutation and those without del(17p)/TP53 mutation but with comorbid conditions rendering them unsuitable for FCR/BR. In addition,

there is an unmet need for patients with del(17p)/TP53 mutation who cannot tolerate ibrutinib (such as those with cardiac risk factors). The company also states there is a high unmet need for treatments that improve progression free survival (PFS) and have potential to achieve undetectable minimal residual disease (see section 4.2.1) in both those with and without del(17p)/TP53 mutation. The ERG clinical advisor agrees with this statement, noting that whilst long-term remission can be achieved with FCR, there is a desire to move to non-chemoimmunotherapy treatments and therefore a need for treatment options for all patients with untreated CLL, regardless of suitability for FCR/BR treatment or mutation status. The company describes how the CLL14 pivotal trial demonstrates that venetoclax plus obinutuzumab (VenG) has the potential to meet the high unmet need in untreated CLL; the ERG reviews this evidence in section 4.2.

Treatment pathway of venetoclax and obinutuzumab (VenG)

The rationale for the treatment combination of VenG is described in the CLL14 trial protocol. Venetoclax is a selective inhibitor of B-cell lymphoma-2, a protein which is overexpressed in approximately 95% of CLL cases.[6] Obinutuzumab is a monoclonal antibody which is directed at the CD20 antigen which is found on most malignant cells of B-cell origin.[7] These different mechanisms were anticipated to improve tumour response in CLL and therefore delay progression and avoid resistance. The treatment combination also allows a chemotherapy-free regimen.

The company presents the current treatment pathway for CLL in the NHS and the positioning of VenG in CS Figure 1. The company states that the anticipated positioning of VenG is:

- For the treatment of previously untreated FCR/BR-unsuitable patients without del(17p)/TP53 mutation
- For the treatment of previously untreated patients with del(17p)/TP53 mutation

The company notes that the anticipated marketing authorisation includes people who would be eligible for FCR/BR (see section 3.1), but considers it likely that in NHS practice VenG will be used in line with the CLL14 study, in which the majority were considered unsuitable for FCR/BR. However, the ERG clinical advisor considers it likely that in practice VenG will be used in younger fitter/patients than those in the CLL14 trial.

3 Critique of company's definition of decision problem

The company's decision problem is largely consistent with the NICE scope, although there are some key differences.

3.1 Population

The population in the company's decision problem is people with untreated CLL with coexisting conditions that make fludarabine and bendamustine based therapy unsuitable.

This is narrower than the NICE scope (people with untreated CLL) but is in line with the population of the pivotal CLL14 trial and with the company's anticipated positioning of VenG in the NHS treatment pathway. However, the anticipated marketing authorisation wording does not specify unsuitability for fludarabine/bendamustine based therapy (see below). The ERG's clinical advisor considered that in UK practice some clinicians are keen to use VenG as a treatment option for patients who are younger/fitter than those in the CLL14 trial, however there is no evidence for use in this population. The NICE scope does not limit by age, but the CS and the anticipated marketing authorisation are limited to adults.

3.2 Intervention

Venetoclax (Venclyxto) with obinutuzumab (VenG). The marketing authorisation is: Venclyxto in combination with obinutuzumab is indicated for the treatment of adult patients with previously untreated CLL. Committee for Medicinal Products for Human Use (CHMP) positive opinion was granted in January 2020, and marketing authorisation in this indication was granted in March 2020.

3.3 Comparators

The company's decision problem includes the comparators according to the following subgroups:

- Without a del(17p)/TP53 mutation: obinutuzumab with chlorambucil
- With a del(17p)/TP53 mutation: ibrutinib

The NICE scope also lists the following comparators. These were excluded by the company with justification as follows:

Without a del(17p)/TP53 mutation

- Fludarabine, cyclophosphamide and rituximab (FCR): the CLL14 trial excludes patients who would normally be suitable for FCR. The evidence submission is for FCR/BR-unsuitable patients only.
- Bendamustine with or without rituximab (BR), for people for whom fludarabine-based therapy is unsuitable: the evidence submission is for FCR/BR-unsuitable patients only.
- Chlorambucil with or without rituximab, for people for whom fludarabine-based therapy is unsuitable: not recommended according to BSH guidelines 2018[2]

With a del(17p)/TP53 mutation

- Idelalisib with rituximab: this has been superseded by treatment with ibrutinib (BSH guidelines 2018[2])

The ERG clinical advisor agrees that it is appropriate to exclude these comparators, but notes that there is a high variability in treatment in current practice. It is also likely that some patients in CLL14 may have been considered suitable for FCR/BR therapy had they been treated routinely in the UK, but it is impossible to quantify this number of patients from summary data.

3.4 Outcomes

Overall survival, progression-free survival, response rate, adverse effects of treatment, health-related quality of life, as per the NICE scope.

3.5 Other relevant factors

The NICE scope and the company's decision problem specify the following subgroups:

- People with untreated CLL with del(17p)/TP53 mutation
- People with untreated CLL for whom fludarabine-based therapy is unsuitable
- People with untreated CLL for whom bendamustine-based therapy is unsuitable.

The first subgroup, people with untreated CLL with del(17p)/TP53 mutation, was a pre-specified subgroup in the CLL14 trial and is considered in the CS. However, the algorithm for identifying the del(17p)/TP53 mutation subgroup differs between the clinical and cost-effectiveness sections, resulting in differing sample sizes (see section 5.2.3 for details). The

rationale for using different algorithms is not adequately justified by the company, although it has little impact on the results.

The other two subgroups, people with untreated CLL for whom fludarabine-based therapy is unsuitable and CLL for whom bendamustine-based therapy is unsuitable, are not addressed separately in the submission, although the CLL14 trial population is considered to be unsuitable for FCR/BR (fludarabine-based/bendamustine-based therapies respectively).

The CS does not include a section on equality considerations. The ERG is not aware of any potential equality considerations for the use of VenG in the UK.

4 CLINICAL EFFECTIVENESS

4.1 Critique of the methods of review

A summary of the ERG's quality assessment of the company's systematic review of clinical evidence is presented in Table 1. The methods of the review were considered appropriate, including searches undertaken and the use of two reviewers for study selection and data extraction, therefore the risk of systematic error in the results of the review is low. The submitted evidence generally reflects the decision problem, although there are differences from the NICE scope in terms of the population and eligible comparators. This is discussed below.

Table 1 Quality assessment of the company's 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 (although additional references were identified by ERG update searches)
3. Is the validity of included studies adequately assessed?	Yes
4. Is sufficient detail of the individual studies presented?	Yes
5. Are the primary studies summarised appropriately?	Yes

4.1.1 Searches

A broad systematic literature review (SLR) was conducted to capture all available evidence for the efficacy, safety and tolerability of all treatments, including those outside of the scope of the company submission, for previously untreated CLL in adults across all populations. The search was not restricted to the two subpopulations considered in the company submission (B1, Table 1).

Searches were conducted in December 2018 with update searches in July 2019. A range of relevant databases were searched: Medline (including Medline In Process), EMBASE, DARE,

NHS-EED, HTA and The Cochrane Library. The ERG note that no trial databases were searched but deem these sources to be appropriate for the identification of relevant literature. Abstracts of major conferences between 2016 and July 2019 were hand searched, as well as the bibliographic references of the systematic and non-systematic reviews found at title-abstract stage. At the full text stage, Letters to editors were searched for RCTs and single-arm studies.

A combination of relevant index and free text terms were used for the main database searches. The search terms for CLL were not as sensitive as the search terms used in a recent Cochrane review [8]. The search was limited by randomised and non-randomised trials. Conference abstracts were excluded from the results, although selected conferences were hand searched. A more limited search was conducted in the Cochrane Library via CENTRAL and Cochrane Database of Systematic Reviews (CDSR).

A total of 150 references were included in the SLR comprised of 56 RCTs (from 36 unique studies) and 94 non-RCT studies (comprising of 80 unique studies). Only 7 references were relevant to the decision problem. 4 references reported one unique study (CLL14) and 3 references reported 3 unique studies (D1.2, Table 8).

CS Appendix Table 7 (CS Appendix D1.2) lists the 150 studies included in the initial clinical SLR but are not of relevance to the decision problem or the submission, but incorrectly includes the eight publications included in the indirect comparison (listed in CS Table 8) (clarification response A28).

4.1.2 Inclusion criteria

The company conducted a broad systematic literature review that aimed to identify studies of all treatments for previously untreated CLL using criteria listed in CS Appendix Table 5 and summarised here:

Population

Established first-line CLL (CLL or b-cell CLL or small lymphocytic lymphoma (SLL)). Inclusion was limited to adults aged ≥18 years (paediatric studies and those where the average age of the population was <18 years were excluded, although the inclusion of individual patients <18 years in an otherwise adult population was allowed). This is in line with the population of CLL14, however it is narrower than the NICE scope which does not limit by age.

Interventions and comparators

A list of twenty interventions were eligible, including all those specified in the NICE scope. Any treatment, no treatment and placebo were eligible comparators.

Outcomes

A list of 21 outcomes were eligible, including most of those specified by the NICE scope (OS, PFS, response rate, and adverse effects). However, HRQoL was not stated. This may mean that relevant studies reporting only HRQoL measures were missed by the company searches, however, a separate SLR of HRQoL was undertaken to inform the economic model.

Study design

Clinical trials and observational studies were eligible.

Other

Only full-text articles and publications in English language were eligible. While this may increase the risk of publication bias, the ERG considers it to be a pragmatic approach.

A two-stage approach was applied to eligibility screening, a flow diagram of study selection is presented in CS Appendix Figure 1. A total of 170 records were initially excluded at full-text review based on the above criteria (a list of these excluded studies and reasons for exclusion was provided in CS Appendix Table 6). The remaining set of 150 records was then limited to studies of:

- VenG and GClb for patients without del(17p)/TP53 mutation
- VenG and ibrutinib for patients with del(17p)/TP53 mutation

A total of 143 records were excluded at this stage, and 7 records were included (reporting 4 unique studies: 1 RCT and 3 non-RCTs).

CS Appendix Table 7 lists all 150 studies identified and included at the initial stage, not just those subsequently excluded because they do not present comparisons of relevance as stated in CS Appendix D.1.2. Reasons for exclusion of the 143 studies subsequently excluded were not given and pdfs of the excluded studies were not provided by the company. The ERG requested

pdfs of any excluded ibrutinib studies regardless of line of therapy (provided in clarification response A14) and checked eligibility. A single-arm study of VenG[9] excluded by the company is summarised by the ERG in section 4.8.

The following comparators specified by the NICE scope were excluded from the company's literature review (see Decision Problem section 3.3 for discussion of this):

- fludarabine, cyclophosphamide and rituximab (FCR)
- bendamustine with or without rituximab (BR), for people for whom fludarabine-based therapy is unsuitable
- chlorambucil with or without rituximab, for people for whom fludarabine-based therapy is unsuitable
- idelalisib with rituximab.

The company has not been explicit about any potential bias in the selection of the studies; however, study selection was undertaken by two independent reviewers.

4.1.3 Critique of data extraction

Data extraction of pre-specified data into extraction tables was undertaken by one reviewer and checked by a second reviewer. The ERG considers the approach to be appropriate.

4.1.4 Quality assessment

The CS provides a risk of bias assessment of CLL14 using the NICE suggested criteria, which include aspects assessing randomisation bias, performance bias and detection bias amongst others. A comparison of the CS and the ERG assessments of the trial is in Table 2.

Overall, the CS considers the CLL14 trial as having a low risk of bias. The ERG generally agrees with this assessment, however notes the performance bias and detection bias inherent in an open label trial. The ERG disagrees with one of the company's judgements as seen in Table 2; the ERG considers that there is a risk of bias due to selective outcome reporting. The ERG also notes that although the arms are balanced with respect to key prognostic factors, there are some differences between groups in baseline comorbidities.

Table 2 CS and ERG risk of bias assessment of CLL14 trial

	CS Response	ERG response
Was randomisation carried out appropriately?	Yes	Yes
Was the concealment of treatment allocation adequate?	Yes	Yes
Were the groups similar at the outset of the study in terms of prognostic factors?	Yes	Yes (in terms of key prognostic factors such as age, sex, mutation status). However, comorbidities were unbalanced (vascular disorders, hypercholesterolaemia, respiratory disorders, psychiatric disorders all >5% difference between groups)
Were the care providers, participants and outcome assessors blind to treatment allocation?	No	No (although Independent Review Committee (IRC) assessments were blinded to allocation)
Were there any unexpected imbalances in drop-outs between groups?	No	No
Is there any evidence to suggest that the authors measured more outcomes than they reported?	No	Yes (published protocol and NCT record lists overall response rate (ORR) at completion of combination treatment assessment, MRD at completion of combination treatment assessment but no reference to these data)
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

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

The CS identified one randomised controlled trial (RCT) (CLL14 trial[10]) which was funded by F. Hoffmann–La Roche and AbbVie Inc. The CLL14 trial compares VenG with GClb in people with known comorbidities that makes them unsuitable for treatment with FCR/BR. In response to clarification question A1, the company confirmed there is no single strategy to confirm suitability for chemotherapy but that participants in the trial had at least one significant co-morbidity that could impact suitability. The trial includes mostly people without del(17p)/TP53 mutation and also a smaller proportion with del(17p)/TP53 mutation. As discussed, the whole trial population is used as evidence for the subpopulation of people without del(17p)/TP53 mutation compared with GClb. The subgroup with del(17p)/TP53 mutation is also used via a naïve indirect comparison with ibrutinib monotherapy – see section 4.4. The evidence for

subpopulation without del(17p)/TP53 mutation therefore includes a proportion of participants with del(17p)/TP53 mutation.

The CS summarises the CLL14 trial in B.2.2 to B.2.7 and further details are reported in CS Appendix D2, D3 and L. In addition, electronic copies of the RCT publication (primary reference Fischer et al 2019[7]) and the confidential Clinical Study Report (CSR)[10] were provided to the ERG, plus a CS addendum and CSR supplement with longer follow-up.

The CLL14 trial is an open label RCT undertaken in people with previously untreated CLL and coexisting medical conditions. The concurrent medical conditions of participants were summarised in B.2.3.6 and Table 10 of the CS and include hypertension, hypercholesterolaemia, cardiac disorders and chronic obstructive pulmonary disease (COPD), see baseline characteristics Table 3 for further details. The ERG clinical expert confirms that these co-existing medical conditions would render the participants unsuitable for treatment with FCR/BR. A concurrent medical condition at baseline was reported in all patients apart from one in the GClb arm (clarification A1).

All participants were aged 18 years or older. They had a life expectancy of more than 6 months, with CLL that required treatment (International Workshop CLL criteria[11]), the presence of coexisting conditions and a total cumulative illness rating scale (CIRS) of >6 (the total score ranges from 0 to 56) or creatinine clearance <70 mL/min. Clinical advice to the ERG is that the latter criterion (CIRS score cut-off of >6 or creatinine clearance <70 ml/min) is not typically used in UK clinical practice to determine lack of suitability for FCR/BR treatment because there is no standard assessment – patients with CLL are assessed individually to include all relevant factors. Our clinical expert considered it likely that some of these patients may have been eligible for FCR/BR treatment in the UK, therefore the trial population may be slightly healthier than the UK population unsuitable for FCR/BR.

The trial had an initial run-in phase where 12 participants received VenG for three cycles to assess safety. Randomisation then proceeded but the CS is unclear whether these 12 participants were included in the RCT. The company confirmed they were not included in the main randomisation phase (clarification response A2). Results from these participants were published[12] although this was not stated in the CS (see section 4.2.1). The trial randomised 216 participants to VenG and 216 participants to GClb. Four participants in the VenG arm and two in the GClb arm did not receive the randomised treatment (reasons provided, CS Appendix Figure

4) but were included in the efficacy analysis (Intention to treat, ITT). At the August 2019 data cut (when all participants had completed 12 cycles of treatment and had a median of 39.6 months follow-up) 177 participants remained in follow-up in the VenG arm and 178 in the GClb arm. There were similar rates and reasons for losses to follow-up across arms.

All participants received intravenous obinutuzumab which was administered for 6 cycles. For cycle 1, 1000 mg was given on days 1 (or 100 mg on day 1 and 900 mg on day 2), day 8 and day 15. Thereafter 1000 mg was given on day 1 of each cycle. Overnight hospitalisation may be required following the first infusion of cycle 1.

For the VenG arm, oral venetoclax was started on day 22 of cycle 1 with an initial ramp-up period (1 week each of 20, 50, 100 and 200 mg daily) then 400 mg daily until the end of cycle 12. The draft summary of product characteristics (SPC) for VenG notes that there is a risk of tumour lysis syndrome (TLS) with venetoclax treatment and describes prophylaxis measures and dose modifications for TLS. People with a high tumour burden and/or reduced renal function have a greater risk of TLS, which occurs when a large number of cancer cells die within a short period, releasing their contents in to the blood. In CLL14, participants deemed at high risk of TLS had the 20 mg and 50 mg doses in hospital (high risk was defined by radiological assessment as any measurable lymph node with the largest diameter ≥ 10 cm or the presence of both ≥ 25 × 109/L absolute lymphocyte count AND any measurable lymph node with the largest diameter ≥ 5 cm but < 10 cm). The dosing of venetoclax was based on the findings from a dose-finding phase I study of venetoclax monotherapy in relapsed or refractory CLL or non-Hodgkin lymphoma.[13]

For the GClb arm, oral chlorambucil (Clb) was administered on days 1 and 15 of each cycle until the end of the 12th cycle at a dose of 0.5 mg/kg. However, this schedule, is not aligned with Clb use in UK clinical practice, where the drug is typically administered over six cycles.

All cycles were 28 days and no cross-overs were permitted.

The CS (B.2.10.2) describes the proportions of participants with dose modification (dose interruption or reduction) of venetoclax, chlorambucil and obinutuzumab during the trial (see Adverse Events section 4.7 for treatment exposure) but there are no details of what the level of dose reductions or alterations were or whether these modifications were defined in the trial protocol. The trial protocol describes permitted dose reductions for adverse events and for TLS,

and the CS reports that at least half of these modifications for one or more of the treatments were for adverse events. The reasons for the remaining dose modifications are not reported. The CS also describes the proportion of participants not reaching the target doses for the three drugs respectively.

The number of patients who discontinued at least one treatment component per treatment arm was 47 (21.8%) for VenG versus 54 (25%) for GClb. The main reasons for discontinuation were adverse events (VenG n=31 versus GClb n=34) or withdrawal of consent (VenG n=9 versus GClb n=11), see CS Appendix Figure 4.

The trial was international and undertaken in 196 sites in 21 countries, including North and South American countries, European countries, and Australia and New Zealand. There were 6 UK sites and 8 UK participants.

Follow-up of the CLL14 trial is currently ongoing; analysis at August 2018 and August 2019 data-cuts have been presented.

Baseline characteristics were similar between groups (Table 3) with the exception of some comorbidities. The CS highlights that there was an imbalance between groups for vascular (specifically hypertension), respiratory, thoracic and mediastinal disorders (in particular COPD and asthma) and for psychiatric disorders (specifically insomnia). All of these were more common in the VenG group. The ERG clinical advisor notes that this may impact on the rate of infective adverse events with greater risk in the VenG group. Additional details of comorbidities were provided in response to clarification A1; these appear balanced between groups.

The numbers of patients with TP53 mutation and del(17p)/TP53 mutation in the CS and CSR are incorrect (101 patients with TP53 mutation were incorrectly categorised as 'unknown'); the company provides an explanation and corrected baseline characteristics and results in clarification response A4. The corrected proportions of patients in each arm with del(17p), TP53 mutation, and del(17p)/TP53 mutation are summarised in Table 3. For completeness, the ERG requested the number of patients in each arm who have both del(17p) and TP53 mutation, but the company did not provide this.

Table 3 summarises corrected key baseline characteristics for the subgroup with del(17p)/TP53 mutation (n=49). There are some imbalances between arms for this subgroup ([redacted]), however for this subgroup it is the comparison with ibrutinib that is relevant to this appraisal (see section 4.5), rather than the comparison with GClb. In clarification response A4 the company provides baseline characteristics and results (PFS and OS) for the subgroup without del(17p)/TP53 mutation (n=368) and the subgroup with these data missing data (n=15); these data have not been summarised in the ERG report.

Table 3 Key baseline characteristics of the CLL14 trial and the del(17p)/TP53 mutation subgroup

% unless stated	Full study population		Subgroup with del(17p)/TP53 mutation	
	VenG (N=216)	GClb (N= 216)	VenG (N= [redacted])	GClb (N= [redacted])
Median (range) Age ≥65 years Age ≥75 years	[redacted] 33.3	[redacted] 36.1	[redacted]	[redacted]
Male sex	67.6	66.2	[redacted]	[redacted]
Median time from diagnosis, months (range)	31.2 (0.4–214.7)	29.2 (0.3–244.8)	[redacted]	[redacted]
High TLS risk	22.2	19.9	[redacted]	[redacted]
Total CIRS score >6	86.1	81.9	[redacted]	[redacted]
Estimated CrCl <70 ml/min, n/N (%)	128/215 (59.5)	118/213 (55.4)	[redacted]	[redacted]
Binet stage A B C	 21.3 35.6 43.1	 20.4 37.0 42.6	[redacted]	[redacted]
ECOG PS 0 1 2 3	 41.2 45.8 12.5 0.5	 47.9 40.5 11.6 0	[redacted]	[redacted]
Cytogenetic subgroup, n/N (%) by the hierarchical model of Döhner et al [14]				
Deletion in 17p	7.9[a]	6.5[a]	[redacted]	[redacted]
TP53 mutational status, Mutated, n/N (%)	[redacted][b]	[redacted]	[redacted]	[redacted]
Del(17p)/TP53 mutation Non-del(17p)/TP53 mutation Missing	[redacted][ab]	[redacted][ab]	[redacted]	[redacted]
IGHV mutational status, mutated	35.2[c]	38.4[c]	[redacted]	[redacted]
Comorbidities % (frequently reported: >30% of patients overall; or imbalanced)				
Vascular disorders	[redacted]	[redacted]	NR	NR
Hypertension	[redacted]	[redacted]	NR	NR
Metabolism and nutrition disorders	[redacted]	[redacted]	NR	NR
Hypercholesterolaemia	[redacted]	[redacted]	NR	NR

Gastrointestinal disorders	[redacted]	[redacted]	NR	NR
Musculoskeletal and connective tissue disorders	[redacted]	[redacted]	NR	NR
Cardiac disorders	[redacted]	[redacted]	[redacted]	[redacted]
Respiratory, thoracic and mediastinal disorders	[redacted]	[redacted]	NR	NR
COPD	[redacted]	[redacted]	NR	NR
Asthma	[redacted]	[redacted]	NR	NR
Psychiatric disorders	[redacted]	[redacted]	NR	NR
Insomnia	[redacted]	[redacted]	NR	NR

CIRS: cumulative illness rating scale; COPD: Chronic Obstructive Pulmonary Disease; CrCl: Creatinine Clearance; ECOG: Eastern Cooperative Oncology Group; NR, not reported; PS: Performance Status; TLS, tumour lysis syndrome. [a] Proportions calculated by ERG using N=216 rather than the N minus missing data as presented in the CS. [b] Values in the CS are incorrect, values here are from clarification response A4 and the CSR corrigendum. [c] CS Appendix Table 46 and CSR reports the proportion with missing data and/or not evaluable and correctly calculates the proportions using the total N in each arm (N=216); the proportions presented in CS Table 9 are different as the company uses the N minus missing data.

4.2.1 Non-RCTs

The CS does not include any non-RCTs. The ERG has identified that the results from the participants included in the run-in to CLL14 were published in a summary paper in 2017.[12] Eleven of the participants completed 12 months of therapy. The publication focus was on safety but response outcomes were also reported (see section 4.8 for participant characteristics and key results).

4.2.2 Ongoing studies

The CS refers to an ongoing study of VenG. The CLL13 trial (NCT02950051) is a multi-centre four-arm RCT that has recruited 926 participants. The clinical trial record states the study is active and that the primary completion date is January 2023. The trial compares standard chemotherapy (FCR or BR) with VenG, venetoclax plus rituximab or triple therapy of VenG plus ibrutinib. The two co-primary outcomes are MRD negativity in peripheral blood and PFS. The population is previously untreated and meets the NICE scope, but does not meet the CS decision problem as participants are physically fit CLL patients (CrCl ≥70ml/min; CIRS ≤ 6; comorbidities excluded). The study is only including participants without del(17p) or TP53 mutation. The ERG have identified one publication summarising the key methods of the CLL13 trial in a conference abstract.[15]

The ERG's searches did not identify any other relevant ongoing studies.

4.3 Description and critique of company's outcome selection

The CS presents all outcomes specified by the NICE scope [overall survival (OS), progression-free survival (PFS), response rates, adverse effects of treatment and health-related quality of life (HRQoL)] from the CLL14 trial. Additional outcomes measured in CLL14 and presented in the CS are minimal residual disease (MRD); duration of response (DOR); event-free survival (EFS) and time to next [anti-leukemic] treatment (TTNT). Safety outcomes were also reported in the CS. These outcomes are summarised in CS Table 7 and the reliability, validity and current use of each outcome in CS Table 8.

Investigator assessed PFS was the primary outcome measure in the trial. PFS was defined as time from randomisation to the first occurrence of progression, relapse, or death from any cause. The International Workshop CLL criteria[11] for PFS were used.

PFS by independent review committee (IRC) was a secondary outcome measure. This used the International Workshop CLL criteria[11] and included at least three experts who were blinded to treatment arm and investigator assessment of response. The investigator assessed and IRC assessed PFS results were similar, with a slightly more favourable hazard ratio (HR) with the IRC assessments (both presented in section 4.4). The latter was used in the CS economic model (section 5.2).

Other secondary outcomes were:

- Overall response rate (ORR), defined as the proportion of participants with a complete response (CR), a complete response with incomplete bone marrow recovery (CRi) or partial response (PR), assessed by the investigator as per International Workshop CLL criteria.[11] ORR was assessed at treatment end and at end of combination treatment assessment (Cycle 7, Day 1 or 28 days after last intravenous infusion), although the latter timepoint was not presented in the CS.
- A composite outcome of CR or CRi at completion of treatment as per International Workshop CLL criteria.[11] The ERG clinical advisor agrees this is a clinically important outcome.
- MRD response rate, defined as the proportion with undetectable MRD in peripheral blood and in bone marrow at completion of treatment and at completion of combination treatment (Cycle 9, Day 1 or 3 months after last IV infusion), although the latter timepoint was not presented in the CS. Undetectable MRD was measured by Allele-

specific oligonucleotide polymerase chain reaction (defined as having < 1 CLL cell per 10,000 leukocytes in peripheral blood or bone marrow). The ERG clinical advisor confirms that undetectable MRD is an important surrogate endpoint, particularly in bone marrow, and that there is a relationship between undetectable MRD and final outcomes in CLL. The CS (B.1.3.3) reports evidence that undetectable MRD leads to improved PFS in CLL and that the European Medicines Agency (EMA) has included undetectable MRD as an intermediate endpoint in recent guidelines for the evaluation of cancer treatments.[16, 17] However, a recent review argues that while MRD status has been shown to be a predictor of PFS and OS following chemo-immunotherapy, data for the relationship between MRD and outcomes following venetoclax are only emerging and need further validation.[18] MRD is not a NICE scoped outcome and has not been used as an outcome in the economic models of previous technology appraisals for CLL. Table 5 of the CS states that MRD outcomes were used in the CS economic evaluation, although there is no explicit description of this in the economic section and the company confirmed at clarifications that this was an error in Table 5 (A18).

- Overall survival was defined as the time from randomisation to death due to any cause. Data were immature at the time of the current analysis. The company clarified that overall survival was used in the economic model, despite CS Table 5 suggesting that it was not (clarification response A18).
- DOR which was defined as time from the first occurrence of a documented overall response to the first occurrence of progression or relapse as determined by the investigator or death from any cause.
- Event-free survival, defined as time between date of randomization and the date of investigator-assessed disease progression/relapse, death, or start of a new anti-leukemic therapy.
- TTNT with an anti-leukaemic agent, which was defined as the time between the date of randomisation and the date of first intake of new anti-leukemic therapy or death from any cause. TTNT was used in the economic model.
- Patient reported outcome measures (PROs) including validated measures of HRQoL (EuroQol 5 dimensions [EQ-5D-3L]; European Organization for Research and Treatment of Cancer Quality of Life Questionnaire Core 30 [EORTC QLQ-C30]) and symptoms (M.D. Anderson Symptom Inventory-CLL [MDASI-CLL]). HRQoL was explored in a scenario in the economic model (clarification response A18).

- Safety, including adverse events and serious adverse events, vital signs, lymphocyte immunophenotyping, premature withdrawals. Adverse effects were included in the economic model.

Exploratory outcomes were: (none of these are discussed further by the ERG)

- MRD (measured with different technologies and using different cut-offs)
- Relationship between blood MRD and PFS
- Relationships between baseline markers and clinical outcomes.

The CS reported all outcomes stated in the protocol and NCT record, although selected timepoints were reported for some outcomes.

4.4 Summary and critique of company approach to statistical analysis and results

4.4.1 Company submissions

The company provided data to the ERG in the following four submissions:

- The original CS and CSR: data-cut August 2018 (28.1 months median follow-up)
- The clarification responses and CSR Corrigendum: data-cut August 2018, correcting errors in the original CS Figure 15 for subgroup analysis of del(17p)/TP53 mutation
- The CS addendum and CSR Supplement: data cut August 2019 (39.6 months median follow-up)
- The CS addendum clarification responses: data-cut August 2019

In the original CS, the company presents multiple outcomes from the CLL14 trial which are consistent with the outcomes specified in their trial protocol.[7] These results are for the whole trial population, unless specified, hence combining patients with and without del(17p)/TP53 mutation. The August 2018 data-cut presented in this submission was not the originally planned primary analysis point, which was scheduled for when 170 PFS events had occurred. Instead, the analyses used the data from the planned interim analysis which was conducted when 65% of the planned PFS events (n=107) had occurred. The interim analysis was originally planned for when 75% of PFS events (n=128) had occurred. The change was based on recommendations from the trial's independent data monitoring committee through a protocol amendment

(version 7) which describes why the interim analysis was performed earlier than originally planned. Hence, these data are very immature and it is not possible to draw conclusions for all of the specified outcomes. This is particularly problematic when it comes to extrapolation performed in the cost-effectiveness section.

In the CS addendum, key outcomes analysed at the August 2019 data-cut are presented. This data cut excluded response measures (which were only assessed at end of treatment) and IRC assessed PFS. A total of [redacted] investigator-assessed PFS events had been observed, still fewer than the number planned for the original primary analysis point. For OS and TTNT, [redacted] and [redacted] events were observed respectively. PROs are not presented in the CS addendum but are available in the updated CSR supplement.

4.4.2 Summary of trial statistics

In the original CLL14 protocol, the company states that analyses would be stratified only by Binet stage. However, the company states that the analyses in their submission are stratified by Binet stage and geographic region (both were stratification factors at randomisation), which is consistent with later versions of the protocol. The reason for this deviation is unclear, but it is not expected to have unduly influenced the results.

The ERG is otherwise satisfied that the analyses based on CLL14 performed by the company are statistically robust and that each analysis was performed on the most relevant population (i.e. ITT or Safety). The trial was well designed and suitably powered to answer its primary hypothesis.

4.4.3 Summary of trial results

A summary of key outcomes from the August 2018 data-cut, or the August 2019 data-cut where available, can be seen in Table 4.

Table 4 Summary of key outcomes

Outcome (95% CI)	VenG (n=216)	GCib (n=216)
Primary outcome: Investigator Assessed PFS (August 2019 data-cut)		
1 year PFS	[redacted]	[redacted]
2 year PFS	88.17 (83.72, 92.61)[a]	64.58 (57.95, 71.20)[a]
3 year PFS	81.9_[redacted]	49.5_[redacted]
Median PFS	Not reached	35.6_[redacted]

	HR 0.31 (0.22, 0.44), p<0.0001	
Secondary outcomes (August 2018 data-cut except where stated)		
1 year PFS, IRC assessed	[redacted] (91.50, 97.71)[a]	[redacted] (87.27, 95.06)[a]
2 year PFS, IRC assessed	88.59 (84.20, 92.98)[a]	63.70 (56.99, 70.42)[a]
	HR 0.33 (0.22, 0.51), p<0.0001	
ORR at EOT	84.7% (79.22, 89.24)[a]	71.3% (64.77, 77.23)[a]
	Difference: 13.43 (5.47, 21.38), p=0.0007[a]	
CR and CRi at EOT	49.5% (42.68, 56.40)[a]	23.1% (17.70, 29.35)[a]
	Difference: 26.39 (17.41, 35.36), p<0.0001[a]	
DOR (August 2019)	HR: [redacted]	
TTNT (August 2019)	HR: 0.51 (0.34, 0.78), p=0.0012	
MRD-Negative blood at EOT	75.5%	35.2%
	Difference: 40.3 (31.5, 49.1), p<0.001	
MRD-Negative bone marrow at EOT	56.9%	17.1%
	Difference: 39.8 (31.3, 48.4), p<0.001	
MRD-Negative blood at 18 months post treatment (August 2019)	47.2	7.4
	[redacted]	
OS (August 2019)	HR 1.03 (0.60, 1.75), p=0.921	
EFS (August 2019)	HR [redacted]	

[a]From FDA report [19]

CR: Complete response: CRi: Complete response with incomplete bone marrow recovery; DOR: Duration of Response; EFS: Event free survival; EOT: End of treatment assessment (3 months); HR: Hazard ratio; IRC: Independent Review Committee; MRD: Minimal residual disease; OS: Overall survival; PFS: progression free survival'; TTNT: Time to next (anti-leukaemic) treatment

4.4.4 Progression-free survival

The primary outcome was investigator assessed progression-free survival (PFS). At the most recent data-cut (August 2019) with a median follow-up of 39.6 months, VenG demonstrated superior efficacy on this outcome, with a hazard ratio of 0.31 (95% CI: 0.22, 0.44; p<0.0001). The company does not present an analysis of whether the hazard ratio, which assumes proportionality of the hazard rates between the two trial arms, is a suitable outcome when reporting this result. However, in their cost-effectiveness section, the company concludes that proportionality is not held. This would suggest that the estimate of the hazard ratio is not an accurate representation of the benefit of VenG on PFS. Despite this concern, the magnitude and statistical significance of the benefit, alongside the visual difference in the treatments on the Kaplan Meier plots (CS Figure 3 and CS addendum Figure 2) mean that the ERG accepts that there is clear and meaningful benefit of VenG over GClb for the primary outcome. Median PFS was not reached in the VenG arm and was 35.6 months (95% CI: 33.7, 40.7) in the GClb arm.

An independent review committee also assessed PFS at the August 2018 data-cut (but not the 2019 data-cut). The results were almost identical to the most recent data cut (investigator-assessed), with a reported hazard ratio of 0.33 (95% CI: 0.22, 0.51; p<0.0001).

4.4.5 Response Rates

Complete and partial response (CR, PR) were assessed in line with the International Workshop on CLL standards[11] and were assessed at the pre-specified end of treatment assessment, conducted 3 months after a patient received their last treatment dose. These outcomes were therefore not updated at the August 2019 data-cut. The company also included in their analysis a complete response with incomplete bone marrow recovery (CRi). The company only present a formal analysis of CR and CRi combined, which demonstrated a 26.4% higher response rate for patients on VenG (95% CI 17.4%, 35.4%; p<0.0001).

A comparison of the PR rates show a lower rate for VenG than GClb (35.2% vs 48.1%), however this may be explained by the higher CR/CRi rate of VenG (49.5% vs 23.1%).

4.4.6 Minimal Residual Disease

Minimal residual disease (MRD, see section 4.3) was measured in both blood and bone marrow of patients at cycle 9 and at 3 months following a patient's last treatment dose. Additional assessments of blood measurements were made at baseline, cycle 7, cycle 12 and every 3 months following end of treatment. There was also an end of treatment (EOT) assessment which occurred 3 months after treatment completion/early termination (Addendum clarification response A3 explained that this was different to the follow-up month 3 assessment). MRD response rate was determined as the number of patients achieving MRD negativity. The main secondary outcome was rate of MRD negativity in blood at the EOT assessment. The company reports that VenG achieved 75.5% MRD negativity compared to 35.2% in the GClb arm at the EOT measurement. At the 18 month post-treatment follow-up (August 2019 data-cut), both arms showed decreased levels of MRD negativity. The company reports 47.2% negativity for VenG and 7.4% for GClb.

A separate assessment of marrow measurements was also undertaken, which at 3 months post-treatment showed lower rates of negativity for both arms than the blood measurements. Here, VenG achieved negativity in 56.9% of patients, compared to 17.1% of patients on GClb. No additional bone marrow assessments were undertaken after this.

The company report that the level of agreement between the two measures at three months post-treatment was high for the VenG arm, however the degree of missing bone marrow measurements makes this difficult to conclude. Whilst agreement of paired measurements for the VenG arm was high at [redacted], it was lower for GClb ([redacted]) suggesting blood measurements of MRD negativity may not be a suitable replacement for bone marrow. The company also performed an analysis of MRD negativity among patients with a CR where the results showed continued benefit of VenG (although note the data presented in CS Table 18 were incorrect, confirmed in clarification response A8). It is also clear that the majority of cases where MRD negativity is achieved are not sustained, for either arm. This could suggest that there is some waning of effect of VenG.

Additionally, the company presents output from an analysis where Kaplan-Meier curves are presented for PFS, but are stratified by treatment arm and MRD negativity status. Whilst no conclusions can be drawn due to a lack of formal hypothesis testing there appears to be a trend demonstrating that MRD negativity from either blood or bone marrow is associated with improved PFS survival. It is apparent that in both arms some patients have a PFS event after having achieved MRD negativity, though it is unclear whether this is disease progression or death.

4.4.7 Overall survival

Overall survival (OS) was a secondary outcome of the CLL14 trial. At the most recent data-cut (August 2019), [redacted] deaths had occurred in both groups (CS Addendum). The hazard ratio for overall survival at the August 2019 data-cut is 1.03 (95% CI: 0.60, 1.75), suggesting that there is no difference in overall survival between VenG and GClb, although there is a degree of uncertainty around the estimate. No assessment of proportional hazards was made alongside the company's presentation of the hazard ratio, but in their cost-effectiveness section the company accept the assumption based on the limited follow-up from CLL14.

4.4.8 Duration of Response

The duration of objective response (DOR) was defined as the time from the first occurrence of a response until a time of disease progression or death. A total of 197 responders to GClb and 200 responders to VenG were included in the analysis. The company described how these numbers were calculated in clarification response A9. There were 60 additional responders in this

analysis (43 from the GClb arm, and 17 from the VenG arm) who had an investigator assessment of response at any time during the study other than for the EOT assessment. These 60 patients were not included in the ORR calculation at EOT. The company again modelled a hazard ratio which assumed proportional hazards without verification of the assumption of proportionality. The ERG believes the assumption is likely to be violated given that the DOR curves for the two arms are identical for the first 10 months before separating. Despite this, at the August 2019 data-cut the stratified hazard ratio of DOR was [redacted]) and the separation of the Kaplan-Meier curves indicates a superior DOR in favour of VenG. This is unsurprising as the responses to VenG were more likely to be a complete response rather than a partial response, compared to GClb.

4.4.9 Event-free survival

Event-free survival (EFS) was defined as the time from randomisation until disease progression or relapse, death or the start of the new anti-leukemic therapy. The company again presents a hazard ratio without consideration of the assumptions made. The analyses were based on [redacted] EFS events on VenG and [redacted] EFS events on GClb. Whilst there was little to distinguish between the Kaplan-Meier curves for the first 12 months of follow-up, the curves did separate beyond this point, in favour of VenG. The hazard ratio of [redacted]) at the August 2019 data-cut indicates that the rate of EFS events was lower on VenG. The events in this analysis are dominated by PFS events, and hence the results are almost identical to those for the PFS outcome.

4.4.10 Time to next treatment

The company defined time to the next anti-leukemic treatment (TTNT) as the time between the date of randomisation and the date of a patient receiving a second line therapy. The original CLL14 trial protocol[7] states "Patients who have not taken new anti-leukemic therapy will be censored at their last assessment prior to the analysis or date of death". However, in their submission it is clear that the company treat death as events in the TTNT analysis, rather than as censored observations, which is consistent with later versions of the protocol.

It is unclear to the ERG how the inclusion of OS events confounds this analysis, in terms of both the magnitude and the statistical significance of the hazard ratio, as the OS events are indistinguishable from the true next treatment events. It is potentially incorrect to include OS

events in the analysis, as the reader infers that that a patient has begun second line therapy when they are actually no longer alive.

Aside from this, the company reports a hazard ratio of 0.51 (95% CI: 0.34, 0.78; p = [redacted]) at the August 2019 data-cut, suggesting that VenG has a significantly lower hazard rate of next treatment or death events than GClb. Evidence from the previous data-cut suggested that the assumption of proportional hazards for TTNT was violated, which is consistent with the crossing Kaplan-Meier curves for the most recent data cut. This casts further doubt on the suitability and interpretability of the hazard ratio reported by the company for this outcome.

The ERG requested an analysis of the TTNT outcome where death events were censored instead of counted as discrete events. In the addendum clarification response B10, the company presented a hazard ratio of [redacted], but without confidence intervals or test of statistical significance. This limited information further supports the conclusion that VenG does delay the TTNT, relative to GClb.

4.4.11 Unreported trial outcomes

Overall response rate at the completion of combination treatment assessment was also reported as a secondary outcome in the CLL14 protocol. This was due at the start of cycle seven or a month after a patient's last intravenous infusion. However, the company have not presented the results for this outcome in their submission, or in the published manuscript of this trial. Whilst the related outcome of response rate at the end of treatment assessment has been reported, it is concerning to see any secondary outcome omitted. The results were in the original CLL14 CSR (data-cut August 2018), which did not indicate a significant difference between treatments. The response rate for GClb was [redacted] and for VenG was [redacted].

In their protocol the company also included a consideration of the MRD response rates in the peripheral blood and in marrow at the completion of combination of treatment assessment, which was due this time at the start of cycle 9 or 3 months after a patient's last intravenous infusion. These results were also omitted from the company submission, but were identified in the original CSR (data-cut August 2018). For peripheral blood at cycle 9, [redacted] of GClb patients were MRD negative compared to [redacted] of VenG patients. For marrow, the proportions were lower for both arms, with GClb achieving MRD negativity in [redacted] of patients, and VenG in [redacted] of patients.

4.4.12 Patient reported outcomes

The company utilised three questionnaires that captured patient quality of life on various scales across the duration of the trial (EQ-5D-3L, MDASI-CLL and EORTC QLQ-C30). The results for each specific assessment can be found in Appendix L of the CS and in more detail in clarification response A10, but in summary, all showed a [redacted] from baseline score, which [redacted]. The company does not report what the baseline values were.

These results [redacted] in terms of change from baseline between the arms of CLL14 at any point in the follow-up. Although the company did not present updated analyses of PROs in their Addendum for the August 2019 data-cut, data in the updated CSR demonstrate that [redacted] This is [redacted] given the observed benefit of VenG, ensuring patients remain progression-free for longer, which is generally associated with a better quality of life. Whilst it is difficult to conclude what may be influencing this [redacted] result, [redacted]

4.4.13 Subgroups

The company presents results of pre-specified subgroup analyses on investigator assessed PFS. The analyses can only be considered exploratory as they were not accounted for in the power calculation and no formal tests of interaction with treatment effect were performed.

There is a discrepancy between the types of subgroups reported in the original CS and the Addendum. Age (<75 vs ≥75), gender (male vs female) and Binet stage (A vs B vs C) were reported only in the original CS (despite the table and text of the Addendum referring to these subgroups). At the August 2018 data-cut, whilst there was a trend of higher relative efficacy of VenG in patients with lower Binet stage, there was consistency across age group and gender (CS Figure 15).

The CS Addendum Figure 8 presents subgroup analyses for TP53 mutation status (mutated, unmutated, unknown) and presence of del(17p), but not the combined subgroup of del(17p)/TP53 mutation presented in CS Figure 15 and defined in the Decision Problem of the

original CS (although the company noted in clarification A4 that there is an error in the del(17p)/TP53 mutation subgroup in CS Figure 15 and provided corrected analyses for the August 2018 data-cut). In response to Addendum Clarification question A1, the company provided investigator assessed PFS by del(17p)/TP53 mutation status at the August 2019 data-cut, see Table 5.

Table 5 Investigator-assessed PFS according to del(17p)/TP53 mutation status, August 2019 data-cut

Subgroup	VenG		GClb		Hazard ratio (95% CI)
	n	Median, months (95% CI)	n	Median, months (95% CI)	
Non-del (17p)/TP53 mutation	184	[redacted]	184	[redacted]	[redacted]
Del(17p)/TP53 mutation	25	[redacted]	24	[redacted]	[redacted]
Undefined mutation status	7	[redacted]	8	[redacted]	[redacted]

NE, not evaluable.

In CS Addendum Figure 8, updated data for IGHV mutational status and cytogenic subgroups are presented. Additional pre-specified subgroups of serum Beta2-microglobulin category, ECOG status, and time from diagnosis to randomisation were also presented. There is consistency of treatment effect across these subgroups.

In the CLL14 protocol, the company specified a further nine subgroups that would be investigated, but are not included in their submission: geographic region, B-symptoms, age (continuous), age (additional categorisations), race, ethnicity, TLS risk, CIRS score and creatinine clearance.

4.5 Critique of trials identified and included in the indirect comparison

Ibrutinib comparator studies

The company identified three studies (Mato 2018,[20] Ahn 2018,[21] ALLIANCE[22]) that could be used to indirectly compare VenG with ibrutinib for people with previously untreated CLL and del(17p)/TP53 mutation. One of these studies, ALLIANCE,[22] was excluded due to the small sample size of the relevant subgroup (n=9). After conducting a feasibility assessment, the company selected Mato 2018 as the preferred study due to its larger sample size, with a secondary comparison using Ahn 2018 presented in CS Appendix D.1.4.

The three ibrutinib studies identified by the company are discussed below.

As noted in section 4.2, the numbers, baseline characteristics and results of patients with TP53 mutation and del(17p)/TP53 mutation in the CS and CSR of CLL14 are incorrect, therefore the company provided corrected baseline characteristics and results in clarification response A4; these are presented in Table 6 below.

Mato 2018

Mato 2018[20] is a retrospective observational cohort study of people with CLL treated with ibrutinib in the front-line setting (see Table 6 for a comparison of study details). Information was obtained from chart review, electronic medical records and related databases. Patients were categorised according to age (< 65 years or ≥ 65 years) and presence or absence of del(17p). These reflect the key inclusion criteria of the pivotal RESONATE-2 RCT of ibrutinib in patients with untreated CLL or small lymphocytic lymphoma (SLL) *without* del(17p) and aged 65 years or over.[23] Mato 2018 also categorised patients separately according to presence or absence of TP53 mutation. The CS is not clear on this point, referring to the relevant subgroup of patients in the Mato 2018 study (n=110) as having del(17p)/TP53 mutation, when in fact their TP53 status is unknown. The ERG notes that the Kaplan-Meier plot for PFS includes 108 del(17p) patients rather than 110, the reason for this is not reported. The ERG also notes that the Mato 2018 whole population contained an additional 8 patients who had TP53 mutations without del(17p) who were not included in the subgroup analysed; these patients are relevant to the current appraisal.

Co-existing conditions with a CIRS score >6 or CrCl < 70ml/min, as required by CLL14, were not reported in Mato 2018, so it is unclear whether the populations were comparable in this respect. Moreover, baseline characteristics were not presented for the subgroup with del(17p). In CS section B.2.9.3, the CS correctly states that the Mato 2018 publication included all ages (whole population n=391, median age 68 years, range 32-96 years, 41% <65 years), although the ages in the relevant del(17p) subgroup are unknown. The CS also states that the CLL14 trial only included patients aged ≥65 years and that '*the inclusion of younger patients in the Mato et al. study could drive the results of the relative comparison to the CLL14 data and generate a trend of ibrutinib superiority*' (CS B.2.9.3). However, the ERG notes that [redacted] of the CLL14 del(17p)/TP53 mutation subgroup were less than 65 years. In response A27 the company clarified that the statement in the CS: 'the CLL14 trial only included patients aged 65 years and above' is incorrect. The median age of the CLL14 whole trial population was 72 years, with [redacted] <65 years.

Overall, the ERG considers that comparability of the Mato 2018 del(17p) subgroup with the CLL14 del(17p)/TP53 mutation subgroup cannot be ascertained. However, based on the characteristics of the whole populations, it is likely that the patients in the Mato 2018 subgroup are younger and fitter. A summary of key results is presented in Table 7.

Ahn 2018

Ahn 2018[21] reports 5-year follow-up of a single arm phase 2 study of ibrutinib in untreated or relapsed/refractory CLL or SLL. Two cohorts are reported: those with age ≥65 years (not relevant as few had del(17p)/TP53 mutation) and those with del(17p)/TP53 mutation. Previous results were published in Farooqui 2015,[24] which was considered in TA429[25] of ibrutinib. In TA429, evidence from Farooqui 2015 was presented for the untreated del(17p) or TP53 mutation population (n=35) but was not used to estimate clinical efficacy as data from the previously treated population were preferred by the company (Committee discussion: *The committee also noted that the single-arm Farooqui et al. (2014) study of ibrutinib presented by the company included a few patients with untreated CLL with a 17p deletion, but that the company did not use this to estimate clinical efficacy. The committee agreed that, in the absence of any evidence, the data from the previously treated population could be taken into account, but recognised this was associated with uncertainty.*)

CS Table 25 summarises baseline characteristics for the del(17p)/TP53 mutation subgroup (n=51) from Ahn 2018, however this includes both untreated (n=35, n=34 in analysis) and relapsed/refractory (n=16) patients with CLL. The company acknowledges this in clarification response A11. CS Appendix Table 9 states that the number of 'previously untreated CLL patients treated with ibrutinib with *TP53* aberrations' is 51, however this is incorrect. Some baseline characteristics for the relevant untreated del(17p)/TP53 mutation subgroup (n=35) are reported by the earlier publication,[24] but there are only three characteristics in common between the studies (age, sex and IGHV mutation) and there is no information on co-existing conditions or CRIS score. The Ahn 2018 subgroup was slightly younger and had fewer men than the CLL14 subgroup. CS page 65 states: '*Farooqui et al. reported on patients with previously untreated CLL patients with TP53 and del(17p), while the Ahn et al. reported on the CLL patients with TP53 aberrations'.* However, the ERG notes that '*TP53 aberrations*' in Ahn 2018 refers to del(17p) or TP53 mutations and, when stratified by treatment status, the untreated subgroup of Ahn 2018 is the same subgroup of patients reported in Farooqui 2015.

CS Appendix D.1.4 states patients without del(17p)/TP53 mutation were excluded from the indirect comparison, reducing the sample size of the population of interest to 24 for the VenG arm from the CLL14 trial and 18 for the ibrutinib arm from Ahn 2018, and that data sources were restricted to elderly patients (65 years and above) only. Clarification response A26 states

that this is incorrect, and that the correct sample size included in the analysis for Ahn 2018 is 34.

Overall, the ERG considers that there are a number of inaccuracies in the description of the Ahn 2018 study by the CS. The comparability of the Ahn 2018 untreated del(17p)/TP53 mutation subgroup with the CLL14 del(17p)/TP53 mutation subgroup cannot be clearly ascertained, although the Ahn 2018 subgroup is younger, has fewer men and is likely to have fewer comorbidities.

ALLIANCE

ALLIANCE[22] is a phase 3 RCT of ibrutinib, ibrutinib + rituximab, and bendamustine + rituximab in people with untreated CLL and age ≥65 years. In the ibrutinib arm, 9 patients had del(17p) and PFS is reported for this subgroup (also 15 patients had TP53 mutation, but results are not presented separately). This study was excluded by the company due to sample size <10. Baseline characteristics are not reported for the subgroup with del(17p) and there is no information on comorbidities or CIRS score. However, as trial participants were randomised to (and therefore suitable for) BR (bendamustine + rituximab) treatment, they are not comparable with the population in CLL14 (unsuitable for BR). This could be considered reasonable justification for exclusion of ALLIANCE from the indirect comparison, however given the lack of appropriate evidence and limitations with the other two studies, the ERG considers that analysis should have been undertaken. This was provided by the company in response to clarification A23.

Feasibility assessment

The ERG agrees that the absence of a common comparator between VenG and ibrutinib precludes an anchored comparison. The company conducted a feasibility assessment to determine the suitability of the available data for conducting an unanchored matching adjusted indirect comparison (MAIC), and concluded that it would not be feasible to conduct a MAIC. After examining the studies, the ERG agrees with this conclusion based on data published in ibrutinib studies. The company stated that unsuccessful attempts were made to contact the authors of the publications (clarification response A25). The ERG contacted the authors of these studies to request baseline data and individual patient data for the relevant subgroups and received a positive response from the ALLIANCE study. However, at the time of writing this report it is unclear whether data will be provided within the timelines of the current appraisal.

ERG summary: Three relevant studies of ibrutinib reported a subgroup of patients with previously untreated CLL and del(17p)/TP53 mutation. The studies did not report comorbidities or CIRS, therefore similarity to CLL14 in this respect could not be ascertained. Baseline characteristics of the relevant subgroups were not reported, therefore MAIC was not possible and comparability with the CLL14 del(17p)/TP53 mutation subgroup is uncertain. There is heterogeneity between these studies and CLL14 in the study designs, eligibility criteria, outcomes and unknown heterogeneity in baseline characteristics. In addition, some of the participants in the CLL14 trial subgroup may have been ineligible for ibrutinib due to cardiac disorders at baseline (clarification A15).

Table 6 Comparison of ibrutinib study details

	CLL14[a]	**Mato 2018**[20]	**Ahn 2018**[21] **(Farooqui 2015**[24]**)**	**ALLIANCE**[22]
Design	RCT	Retrospective observational cohort study	Single arm phase 2 study	RCT
Eligibility criteria	• Age ≥18 years • Life expectancy > 6 months • Previously untreated CLL • Total CIRS score >6 or CrCl <70 mL/min • CrCl ≥ 30 ml/min	• Previously untreated CLL • Treated with ibrutinib	• CLL or SLL • del(17p) or TP53 mutation for the TP53 cohort [*or* age ≤65 for elderly cohort] • ECOG PS ≤2 • Creatinine <2.0 mg/dL or CrCL 50 mL/min or less • Previously untreated CLL or R/R CLL	• Age ≥65 years • Previously untreated CLL • Intermediate or high-risk Rai stage CLL • ECOG PS ≤2
Relevant subgroup	**del(17p)/TP53 mutation: 25** del(17p): 17 TP53 mutation: 23	**del(17p): 110 (108 in analysis)** TP53 mutation: 44 Both del(17p) and TP53 mutation: 35[b] TP53 mutations without del(17p): 8	**del(17p)/TP53 mutation: 35 (34 in analysis) untreated**	**del(17p): 9 (untreated)** TP53 mutation: 15
Baseline characteristics reported by more than one study				
	VenG (n=25)	Ibrutinib (n=108)	Ibrutinib (n=35)	Ibrutinib (n=9)
Age	[redacted]	Not reported	median 62 (range 33‑82)	Not reported
% male	[redacted]	Not reported	66%	Not reported
IGHV un-mutated	[redacted]	Not reported	63%	Not reported

[a]baselines as reported in clarification response A4. [b] States 34 in results section.

Table 7 Key results in subgroup with untreated del(17p) and / or TP53 mutation (studies in indirect comparisons)

	CLL14 VenG (n=25)[a]	Mato 2018[20] Ibrutinib (n=108)	Ahn 2018[21] (Farooqui 2015[24]) Ibrutinib (n=34)	ALLIANCE[22] Ibrutinib (n=9)
ORR	NR	82.3% Both del(17p) and TP53 mutation (n=34): 91%	NR	NR
CR	NR	Clinical CR 21.2%	12%[b]	NR
PR	NR	43.5%	70% [b]	NR
PR with lymphocytosis	NR	17.6%	15% [b]	NR
SD	NR	13.0%	-	NR
PD	NR	4.6%	3% [b]	NR
Discontinuation rate	NR	33%	NR	NR
Mean time to discontinuation	NR	6.25 months	NR	NR
PFS	2 year: [redacted]%[a]	1 year: 87%	5 year: 74.4% (95% CI 60.2, 92.1)	2 year: 75% [c]
OS	1 year: [redacted][a]	n=103 1 year: 89%	5 year: 85.3% (95% CI 74.2, 98.1)	NR

[a]Results from clarification response A4 ([redacted]), August 2018 data-cut. [b]best response at 24 months follow-up. [c] estimated from Kaplan-Meier curve. NR, not reported.

4.6 Critique of the indirect comparison

In the absence of direct evidence, the company sought to perform an indirect comparison of VenG with ibrutinib for patients with del(17p)/TP53 mutation. In the CLL14 trial, there were just 25 patients with del(17p)/TP53 mutation who experienced [redacted] PFS events (August 2019 data-cut). The numbers of relevant patients in the comparator trials are 108 (Mato 2018.[20]), 34 (Ahn 2018 [21]) and 9 (Woyach 2018[22]) meaning any comparison would likely be considerably underpowered. When combined with previously discussed issues of heterogeneity, any comparison made will be extremely limited in its validity.

The company concluded that a MAIC was not suitable given the lower number of patients that would be eligible from CLL14 based on matching to the ibrutinib trial inclusion/exclusion criteria, even before matching to specific covariates. A MAIC is useful when you have access to patient level data from one trial, and apply weights to each patient such that the distribution of key population level variables match that of an arm of a target trial. Matching typically reduces the final overall sample size by reducing the weight of certain participants who do not match well to the target population. Hence, the ERG accept that a MAIC analysis would not be ideal, but

requested that the company attempt it given the heterogeneity present in the studies that may bias a naïve treatment comparison.

In their original submission, the company performed a naïve treatment comparison, which did not make any covariate adjustment and assumed the ibrutinib trials contained patients homogenous to those in the del(17p)/TP53 VenG population of CLL14. Clearly, this contains a number of significant risks, and the ERG advises that no conclusion should be drawn from such an analysis. The company perform the comparison utilising hazard ratios, but without any assessment of whether these are a suitable scale to compare the treatments. Also, digitising graphs to obtain patient level data that is representative of the trial data, but not necessarily identical, which is another source of uncertainty within these comparisons. Updated analyses using the August-2019 data cut are presented in the CS addendum.

The company's first comparison is to Mato 2018. Fitting a Cox proportional hazard model to the data produced a PFS hazard ratio of 0.660 (95% CI: 0.270, 1.615; p = 0.363). The wide confidence intervals suggest no conclusion can be drawn even if the assumptions of the analysis were valid. Furthermore, it is likely that this effect size is capturing a combination of treatment effect and differences in prognostic factors.

Fitting another Cox model to the OS data produced a hazard ratio of 0.841 (95% CI: 0.301, 2.352; p=0.741). Again, this analysis did not provide any useful information to meaningfully estimate a relative effect on OS between VenG and ibrutinib in del(17p)/TP53 mutation patients.

A naïve comparison to the most relevant patients from the study of Ahn 2018 produced a hazard ratio of [redacted] (95% CI: [redacted], [redacted]; p=[redacted]) for PFS, favouring ibrutinib. The OS data yielded a hazard ratio of [redacted] (95% CI: [redacted], [redacted]; p=[redacted]), also in favour of ibrutinib but not significantly.

The ERG requested additional analyses which pooled the recreated data for the ibrutinib patients prior to obtaining hazard ratios. The results were similar with the single study analyses, with a PFS hazard ratio of [redacted] (95% CI: [redacted], [redacted]; p=[redacted]) and an OS hazard ratio of [redacted] (95% CI: [redacted], [redacted]; p=[redacted]) (CS addendum clarification response B7), however it is

not possible to conclude whether these results could be considered more reliable than the others.

In summary, these indirect comparisons are inadequate for providing any meaningful information on the comparison of VenG and ibrutinib for either PFS or OS in patients with del(17p) or TP53 mutation.

4.7 Adverse events

The safety-evaluable population was used for the safety analysis, this population was defined as participants who received at least one dose of any study treatment. There were 212 participants in the safety analysis population for VenG and 214 for GClb. The CS presents only adverse events that were defined as treatment-emergent adverse events (TEAEs), these were events not present at the start of study treatment or an event that was already present which worsened with study treatment.

Treatment exposure rates were presented in CS Section B.2.10.2 and are summarised by the ERG in Table 8. Data were not available for some of these categories and it is therefore difficult to compare between treatments across all factors. The median dose intensity rate of venetoclax in the VenG arm was [redacted] and the median dose intensity of chlorambucil in the GClb arm was 95.4% (range 4-111%). In both arms the median dose intensity of obinutuzumab was 100% (range 0-111%). The proportion of participants with drug interruption or reductions ranged from [redacted] for obinutuzumab to 43.3% for venetoclax in the VenG arm, and between 26.9% on chlorambucil and [redacted] on obinutuzumab in the GClb arm.

Table 8 Summary of treatment exposure rates in CLL14 (safety-evaluable population)

	VenG (n=212)	**GClb (n=214)**
Completion of treatment		
Both agents Single agent period	[redacted] Venetoclax: [redacted] Obinutuzumab: NR	NR Chlorambucil: NR Obinutuzumab: NR
Number of cycles, median (range) per agent	Venetoclax: NR Obinutuzumab[redacted]	Chlorambucil: [redacted] Obinutuzumab[redacted]
Median duration of exposure from first dose, months (range)	[redacted]	NR

	VenG (n=212)	GClb (n=214)
Median (range) dose intensity per agent (for venetoclax this is after reaching target dose)	Venetoclax: [redacted] Obinutuzumab: 100% (0-111%)	Chlorambucil: 95.4% (4-111%) Obinutuzumab: 100% (0-111%)
Median total cumulative dose per agent	Venetoclax: NR Obinutuzumab: [redacted] mg	Chlorambucil: NR Obinutuzumab: [redacted] mg
Reached target dose single agent, n/N	Venetoclax: [redacted] Obinutuzumab: NR	Chlorambucil: NR Obinutuzumab: NR
Dose modification (interruption or reduction) rate in those reaching target dose, % per agent	Venetoclax: 43.3% Obinutuzumab: [redacted]	Chlorambucil: 26.9% Obinutuzumab: [redacted]

[a] [redacted] did not reach the 400mg target dose for a variety of reasons, including AEs and withdrawal of consent. NR: Not reported (in the CS or CSR)

4.7.1 Overview of treatment-emergent adverse events

Table 9 provides a summary overview of the key rates of adverse events in CLL14 and these are described in more detail below. Adverse events (AEs) were collected until 28 days post-treatment, and grade 3-4 adverse events (other than grade 3-4 infections, which were reported for 2 years after the last dose) were collected until 6 months post treatment, therefore these were not updated in the August 2019 data-cut (although there are minor differences in the update due to data cleaning and administrative updates). Serious adverse events (SAEs) and fatal AEs were updated at the August 2019 data-cut.

TEAEs were experienced in [redacted]% of participants in the VenG arm and [redacted]% of participants in the GClb arm.

There are minor differences between data reported in the original CS (and trial publication,[7] VenG 78.8% vs GClb 76.6%) and the updated CSR supplement (VenG [redacted] vs GClb [redacted]) due to data cleaning or administrative updates. Corrected data from the CSR supplement are presented below where available. CS Addendum Table 12 also reports grade 3-4 adverse events 'at greatest intensity' of [redacted] with VenG and [redacted] with GClb.

At the August 2019 data-cut, SAEs were experienced in [redacted]% of participants in the VenG arm and [redacted]% of participants in the GClb arm. The all-cause death rate during the trial was similar between groups, however, deaths due to adverse events were higher in the VenG arm ([redacted]% vs [redacted]%) (Table 9). Treatment discontinuation rates (any treatment) were similar between arms.

Table 9 Summary of adverse events in CLL14, safety-evaluable population

%	VenG (n=212)	GClb (n=214)
Any treatment-emergent AE	[redacted]	[redacted]
Treatment-related grade 3-4 AE	[redacted]	[redacted]
Any treatment discontinuation for TEAE	14.6	15.9
At least one SAE (August 2019)	[redacted]	[redacted]
Death (any cause) (August 2019)	[redacted]	[redacted][a]
Death related to AE (August 2019)	[redacted]	[redacted]

[a] Excludes one participant who died prior to randomisation
AE: Adverse event, SAE: Serious Adverse event; TEAE: treatment-emergent adverse event.

4.7.2 Grade 3-4 adverse events

National Cancer Institute Common Terminology Criteria for Adverse Events were used to assess the severity of AEs.[7] The CS reports the grade 3-4 TEAEs that had a difference of at least 2% between treatment arms (coded using MedDRA v21.0) (CS Table 28) and those with an incidence of at least 1% in either arm because these were used in the CS economic evaluation (CS Table 29). Grade 3-4 TEAEs with at least 2% greater incidence in the VenG arm were neutropenia ([redacted] VenG versus [redacted]% GClb); hyperglycaemia (3.8% VenG versus 1.4% GClb); diarrhoea ([redacted] VenG versus [redacted] GClb August 2019) and hypertension ([redacted] VenG versus [redacted] GClb). Grade 3-4 leukopenia was more commonly experienced in the GClb group compared with the VenG group ([redacted] VenG versus [redacted] GClb). Neutropenia and diarrhoea are known adverse drug reactions related with venetoclax.

Key grade 3-4 TEAEs used in the CS economic model can be seen in Table 10; rates were higher in the VenG group for asthenia ([redacted] VenG) and dyspnoea ([redacted] VenG) versus GClb (both events [redacted]); febrile neutropenia ([redacted] VenG versus [redacted] GClb) and sepsis ([redacted] versus [redacted] GClb). The ERG notes that sepsis has a difference of greater than 2% between groups.

Table 10 Grade 3-4 TEAEs used in the CS economic evaluation (updated CSR)

%	VenG (n=212)	GClb (n=214)
Asthenia	[redacted]	[redacted]
Diarrhoea	[redacted]	[redacted]
Dyspnoea	[redacted]	[redacted]
Febrile neutropenia	[redacted]	[redacted]
Infusion related reaction	[redacted]	[redacted]
Leukopenia	[redacted]	[redacted]
Neutropenia	[redacted]	[redacted]
Pneumonia	[redacted]	[redacted]

Sepsis	[redacted]	[redacted]
Thrombocytopenia	[redacted]	[redacted]

4.7.3 Serious adverse events and deaths

SAEs were defined as any adverse event that is fatal, life threatening, requires or prolongs hospital stay, results in persistent or significant disability/incapacity, any congenital anomaly/birth defect in a neonate/infant born to a mother exposed to the study drug or any other significant medical event.[7] As seen in Table 9 above, SAEs were experienced in more participants ([redacted]%) in the VenG arm than in the GClb arm ([redacted]%) at the August 2019 data-cut. The company did not provide details of SAEs for the treatment and post-treatment phases of the trial combined. The ERG has therefore presented SAEs (experienced by at least 1% of participants in at least one arm of the CLL14 trial) reported in the post-treatment period (from CS Addendum Table 13) alongside those reported during the treatment phase of the trial (shown in Table 11). The most frequently reported SAEs in the VenG arm at the August 2018 data cut were febrile neutropenia, pneumonia, infusion-related reaction and pyrexia. These were also the most frequently reported SAEs in the GClb arm and of these only febrile neutropenia (higher for VenG) and infusion-related reactions (higher for GClb) had a ≥1% difference between groups. Overall there were no SAEs that were experienced ≥2% more in one of the groups.

Table 11 Summary of SAEs with ≥1% incidence in either treatment group

	August 2018 data cut		**Post treatment period, August 2019 data cut**	
%	**VenG (N=212)**	**GClb (N=214)**	**VenG (N=202)**	**GClb (N=208)**
Pneumonia	4.7	4.2	[redacted]	[redacted]
Sepsis	2.8	0.9	[redacted]	[redacted]
Cellulitis	1.4	0	[redacted]	[redacted]
Infusion-related reaction	4.2	6.1	[redacted]	[redacted]
Febrile neutropenia	5.2	3.7	[redacted]	[redacted]
Thrombocytopenia	0.9	2.3	[redacted]	[redacted]
Neutropenia	1.4	0.5	[redacted]	[redacted]
Squamous cell carcinoma	0.9	1.4	[redacted]	[redacted]
Pyrexia	3.8	3.3	[redacted]	[redacted]
COPD	1.4	0.9	[redacted]	[redacted]
Atrial Fibrillation	0.5	1.4	[redacted]	[redacted]
Cardiac Failure	1.4	0.5	[redacted]	[redacted]
Myocardial infarction	0.5	1.4	[redacted]	[redacted]
Tumour lysis syndrome	0.5	1.9	[redacted]	[redacted]
Aspartate aminotransferase increased	0	1.9	[redacted]	[redacted]
Alanine aminotransferase increased	0	1.4	[redacted]	[redacted]

COPD: Chronic Obstructive Pulmonary Disease. Additional SAEs occurring during the post-treatment period: respiratory tract infection, prostate cancer, cerebral ischaemia, dehydration, hypertension, vertigo (all VenG 1% vs GClb <1%, CS Addendum Table 13).

As described above, there were more deaths related to TEAEs with VenG (see Table 9). Sepsis / septic shock was the most frequently reported TEAE leading to death (participants in the VenG arm and in the GClb arm, August 2019 data-cut). The CS states that a causal association with venetoclax and death was unlikely because of the long latency period from the last dose of study drug (of deaths assessed in August 2018, 11 of the VenG arm died 29 days or more after last study drug), pre-existing medical conditions and concomitant comorbidities. The ERG notes that four of the eight participants in the GClb arm who had died as a result of TEAEs at the August 2018 data cut died in the post-treatment period and that participants in both arms had pre-existing medical conditions and concomitant comorbidities. Two of the deaths in the VenG arm at the August 2018 data cut were attributed a causal relationship to obinutuzumab by the investigator.

4.7.4 Adverse events of any grade

Table 12 provides a summary of specific adverse events with ≥10% incidence in either treatment group (reproduced from the Fischer publication of the CLL14 trial,[7] data-cut August 2018) for context. Events with a 5% or greater difference between groups are in bold.

Table 12 Overview of AEs with incidence of ≥10% in either group at August 2018

%	VenG (N=212)	GClb (N=214)
Neutropenia	57.5	57.0
Thrombocytopenia	24.1	23.4
Anaemia	16.5	18.7
Infusion-related reaction	**44.8**	**51.4**
Diarrhoea	**27.8**	**15.0**
Nausea	18.9	21.5
Constipation	13.2	8.9
Pyrexia	**22.6**	**15.4**
Fatigue	15.1	14.0
Cough	16.0	11.7
Headache	11.3	9.8

4.7.5 Tumour lysis syndrome

Tumour lysis syndrome is an important consideration in the treatment of CLL (section 4.2). At the August 2018 data-cut, TLS was reported in three VenG treated participants and in five GClb treated participants. All cases in the VenG arm occurred during treatment with obinutuzumab

and before treatment with venetoclax and none met the Howard criteria for clinical TLS; that is the presence of specific electrolyte changes and clinical manifestations.

4.8 Additional work on clinical effectiveness undertaken by the ERG

Updated searches for published and ongoing studies were undertaken by the ERG.

Eight new publications relevant to the submission were identified: two VenG studies[9, 12] (see section 4.8.1), one GClb study[26], one abstract linked to the ongoing CLL13 study[15] (see section 4.2.2) and four abstracts linked to CLL14.[27-30] The CLL14 abstracts were checked for additional data but none were identified. No ongoing studies were identified.

The two additional studies of VenG[9, 12] identified by the ERG are summarised here. In addition, the ERG has summarised the CLL11 trial[31], which is referred to by the company for external validation of the GClb arm, and the ERIC real-world study of GClb.[26]

4.8.1 Additional VenG studies

The results from the participants included in the run-in to CLL14 were published in a summary paper in 2017.[12] This was not reported by the CS.

Thirteen previously untreated CLL patients received VenG. The dose regimen was the same as for participants of CLL14 (section 4.2). Baseline characteristics of these participants are in Table 13.

Table 13 Baseline characteristics of CLL14 run-in participants

% unless stated	N=13
Median (range) Age ≥70 years	75.0 (59-88) 84.6
Male sex	61.5
Total CIRS score >6	76.9
Estimated CrCl <70 ml/min	76.9
Binet stage A B C	 15.4 23.1 61.5
Deletion in 17p	2/8 (25.0)
TP53 mutational status, Mutated, n/N (%)	2/8 (25.0)
TP53 deleted and/or mutated, n/N (%)	2/8 (25.0)
IGHV mutational status, mutated, n/N (%)	1/7 (14.3)

Eleven of the participants had completed 12 months of therapy at the time of the data cut. One of the non-completers developed a grade 4 infusion related reaction at the first obinutuzumab dose and one chose to discontinue at cycle 8.

Median follow-up was 15 months. Response rates and MRD negative rates three months after the end of treatment are summarised in Table 14; complete response was seen in 58% and partial response in 42%. All participants experienced at least one adverse event (Table 14). Grade 3-4 AEs were experienced in 83.3%; these included neutropenia (58.3%), febrile neutropenia (25.0%), TLS (16.7%) and infusion-related reactions (8.3%). The authors concluded that VenG could be safely administered to patients with comorbidities and at risk of TLS due to renal impairment.

Table 14 Available efficacy and key adverse event data CLL14 run-in

Outcome at EOT, %	**VenG (n=12)**
ORR	100
CR	58
PR	42
PD	0
MRD negative (peripheral blood)	91.7
MRD negative (bone marrow)	5/7 (71.4)
Adverse events, %	
Any AE	100
At least one grade 3/4 AE	83.3
Death	0

AE: Adverse event; CR: Complete response; EOT: End of treatment (3 months after completion of last cycle); MRD: Minimal Residual Disease; ORR: Overall response rate; PD: Progressive disease; PR: Partial Response.

The CS included a non-RCT of VenG by Flinn 2019[9] in the initial clinical SLR but it was subsequently excluded because it was considered not relevant to the decision problem (CS Appendix Table 7). This phase 1b single-arm study of VenG included two cohorts, those who were treatment naïve and those with relapsed/refractory CLL. Results for a subgroup of treatment naïve participants with del(17p)/TP53 mutation were also reported (n=5). The cohort with no prior treatment (and the subgroup with del(17p)/TP53 mutation) meet the wider NICE scope, but the ERG agrees the population does not meet the company's decision problem as the participants were not considered unfit. However, the ERG has summarised the limited results for efficacy and safety for context.

The study was a dose finding and safety expansion study. As part of the dose finding phase participants either received venetoclax first or obinutuzumab first during cycle 1 to reduce TLS risk. Thirty-two CLL participants with no previous treatment were administered VenG for 6 cycles and then venetoclax was given as a monotherapy until disease progression, unacceptable toxicity or completion of 1-year treatment. Twelve participants were enrolled during the dose finding phase and 20 during the safety expansion, but all received venetoclax 400mg. The study was performed in 11 centres including at least one from the UK. The key baseline characteristics of the treatment naïve cohort are summarised in Table 15.

Table 15 Baseline characteristics of participants with no previous treatment for CLL from Flinn 2019[9]

% unless stated	**N=32**
Median (range)	63 (47-73)
Male sex	63
Estimated CrCl <70 ml/min	29
ECOG PS 0 1	 50 50
TP53 mutation	16[a]
Del(17p)/TP53 mutation	16[a]
IGHV unmutated	50[a]

[a]Calculated by ERG using the total sample N as the denominator. The publication reports the proportion using the denominator as N minus missing data.

CrCl: Creatinine clearance; ECOG: Eastern Cooperative Oncology Group; PS: Performance status

Median follow-up for the treatment naïve cohort was 26.7 months (range, 16-39 months). Key results of efficacy and adverse events can be seen in Table 16. Results for 24-month PFS were in the same region as the VenG population in CLL14 but ORR and CR/CRi rates were better in this cohort. Undetectable MRD was an exploratory outcome. Rates of undetectable MRD in peripheral blood were 91% at least 3 months after the last obinutuzumab dose and 72% after median 4.4 months from the last venetoclax dose. The 'best response' rate of undetectable MRD in bone marrow was 78%. However, from patient level data it was apparent that of the 25 (78%) patients who achieved bone marrow negativity at least once, 15 of these (60%) later had either a positive blood or bone marrow test, suggesting the negativity was not sustained in the majority of patients.

Adverse event rates were similar between this cohort and the VenG arm of the CLL14 trial.

Table 16 Available efficacy and key adverse event data from Flinn 2019[9]

Outcome at EOT, %	**N=32**
PFS at 24 months	90.6 (95% CI 80.5-100%)
ORR (best response)	100 (95% CI 89-100)
CR/CRi	78 (95% CI 60-91)
PR	22
PD	12.5
MRD negative (peripheral blood) >3 months after last obinutuzumab treatment Median 12 months from last obinutuzumab treatment Median 4.4 months from last venetoclax treatment	 91 78 72
MRD negative (bone marrow) Best response achieved	 78
Adverse events, %	
Any AE	100
At least one grade 3/4 AE	78
Any SAE	34
Venetoclax discontinuation due to AE	3
Death	0

AE: Adverse event; CR/CRi: Complete response / Complete response with incomplete bone marrow recovery; MRD: Minimal Residual Disease; ORR: Overall Response Rate; PD: Progressive Disease; PFS: Progression-free Survival; PR: Partial Response; SAE: Serious Adverse event

Some efficacy data for the del(17p)/TP53 mutated subgroup were also reported and are summarised in Table 17, although the ERG note the small sample size rendering comparison unreliable.

Table 17 Efficacy data for the del (17p) / TP53 mutation subgroup from Flinn 2019[9]

Outcome at EOT, %	**N=5**
ORR (best response)	100
CR/Cri	60
PR	40

CR/CRi: Complete response / Complete response with incomplete bone marrow recovery; ORR: Overall Response Rate; PR: Partial Response

4.8.2 GClb studies

CLL11 trial

The company refers widely to the CLL11 trial (Goede 2014[31]) for external validation of the GClb arm. CLL11 was the pivotal trial in TA343 of Obinutuzumab in combination with chlorambucil for untreated CLL.[32]

The 3-arm trial compared chlorambucil, GClb and rituximab with chlorambucil in people with previously untreated CLL and comorbidities reflected in either ≥6 on CIRS or CrCl 30-69

ml/min, therefore the target population is similar between the two trials. However, while the CS states that CLL14 does not include patients who would receive FCR or BR in clinical practice (CS Table 2), CLL11 included people unsuitable for fludarabine-based treatment; some of these were suitable for bendamustine treatment. NICE recommends GClb only for the subgroup for whom bendamustine-based therapy is not suitable.[32] This subgroup is more relevant to the CLL14 trial, however results have not been not published.

Key baseline characteristics such as age, sex, Binet stage, CIRS >6, IGHV mutation status and del(17p) status were similar (difference <10%) between GClb arms in the CLL11 and CLL14 studies (although TP53 mutation status was not reported by CLL11). There were higher rates of cardiac and respiratory comorbidities in the CLL11 GClb arms compared with CLL14 at baseline.

The planned dose of GClb was similar between CLL11 and CLL14, except that chlorambucil was given for six cycles in CLL11 compared with twelve cycles in CLL14. The ERG for TA343 noted that the dose of chlorambucil used in CLL11 (about 70 mg) was lower than that generally used in clinical practice in England (about 120 mg).[32] The median dose intensity for chlorambucil in the GClb arm of CLL14 is reported to be 95.4% (range: 4%–111%), but it is not clear how this relates to clinical practice.

Overall, the CLL14 and CLL11 trials are similar in most aspects, although there are some key differences.

ERIC study

A recent retrospective multi-centre study (ERIC[26]) assessing the use of obinutuzumab with or without chlorambucil in 437 treatment-naïve patients from Europe, Israel, Canada and Argentina in a 'real-world' setting was identified by the ERG. The majority of patients received GClb (n=408). Those with del(17p) or TP53 mutations were excluded from the study.

The target population is similar between CLL14 and ERIC, in that the participants are described as 'unfit'. However, there are a few differences in baseline characteristics between CLL14 and the GClb cohort of ERIC. Although the median age in ERIC is similar to CLL14, the minimum and maximum ages are higher suggesting a slightly older participant group, and median time from diagnosis for the whole cohort in ERIC is longer (although it is unclear if this is measured to the

same point in the treatment history). There are also slightly fewer men in ERIC. The CLL14 trial required participants to have a total CIRS score > 6 or CrCL <70 ml/min. In ERIC, the proportion with CIRS >6 is lower than in CLL14, but the proportion with CrCL <70 ml/min is higher. Binet stages are generally similar in both studies. The ERIC study doesn't report the presence of cardiac or respiratory comorbidities and overall it is unclear whether the ERIC population is less or more fit than the population of CLL14.

Treatment with GClb was for 6 cycles in ERIC, in line with the regimen used in the CLL11 study (see Section 4.8.2). This is different to CLL14 where 12 cycles were used. The study periods in the two studies are similar: ERIC included patients who were treated during 2014-2019, while CLL14 recruited patients 2015-2016, but median follow-up was shorter in ERIC (14.1 months) than CLL14 (39.6 months).

4.9 Conclusions of the clinical effectiveness section

The evidence for the effectiveness of VenG compared with GClb for people with untreated CLL with coexisting conditions which make FCR/BR based therapy unsuitable, is from a good quality RCT. Improved PFS, CR and DOR were found with VenG, but no difference in OS was observed. Despite the submission of a more recent data-cut by the company, data remain immature for some key outcomes. The ERG noted concerns regarding the generalisability of the CLL14 trial to the UK population.

There is no head-to-head comparison between VenG and ibrutinib for the subgroup with del(17p)/TP53 mutation. Naïve indirect comparisons with three ibrutinib studies produced hazard ratios that suggested that VenG was inferior to ibrutinib for PFS and OS, however the concerns around the suitability of the comparison and the width of the confidence intervals mean that no conclusion of superiority can be drawn. The company used two different methods in the CS to identify the subgroup with del(17p)/TP53 mutation, resulting in different sample sizes between the clinical and cost-effectiveness sections. The rationale for using different algorithms is not adequately justified by the company, although it has little impact on the results.

5 COST-EFFECTIVENESS

This chapter reviews and appraises the submitted evidence on the cost-effectiveness of VenG for untreated CLL. Section 5.1 gives the ERG's critique of the company's systematic reviews. Section 5.2 provides a summary and critique of economic aspects of the CS. Section 5.3 presents the ERG's preferred base case estimates and additional work carried out by the ERG. Section 5.4 provides the conclusions of the cost-effectiveness section and section 5.5 looks at the impact on the incremental cost-effectiveness ratio (ICER) of additional analyses.

The CS was received on 29 October 2019. This submission was based on CLL14 trial data available from the August 2018 data cut (28.1 months median follow-up time from randomisation) and is referred to as the original CS. In addition, as explained in Section 4.4.1, the company submitted additional information to the ERG as follows:

- Responses to ERG's request for clarifications on the original CS. This is referred to as the 'original CS clarification responses'.
- Updated analysis, submitted as an addendum, taking into account a newer data cut-off (August 2019, 39.6 months median follow-up time from randomisation). This is referred to as the 'CS addendum'.
- Responses to ERG's request for clarifications on the CS addendum. This is referred to as the 'CS addendum clarification responses'.

Similarly, the original economic model was received on 29 October 2019 and it was based on the CLL14 August 2018 data-cut. Additional models were submitted subsequently, in response to clarification queries and availability of newer data (see Section 5.2 below).

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

5.1.1 Objective of cost-effectiveness review

The company carried out and reported a SLR of cost-effectiveness evidence. The aim of the review was to identify studies within the literature on previously untreated CLL that reported (i) relevant economic evaluations, (ii) health-related quality of life (HRQoL), and (iii) costs and use of health care resources. The scope of the search is broader than final NICE scope as it is not restricted to any specified treatment or subpopulation. Literature searches were initially carried out in December 2018 and were subsequently updated in July 2019.

Searches were conducted in a range of sources, including key electronic bibliographic databases (MEDLINE, Embase, EconLit, DARE, NHS-EED, HTA and Cochrane Library Databases), conference abstract books, HTA websites, databases and reference lists of relevant published systematic and non-systematic reviews. The ERG deems these sources to be appropriate for the identification of relevant literature. Search terms were split into key 'topics' (facets) including treatment setting, condition, cost-effectiveness, health care resource use and costs and HRQoL and terms relating to each topic (including synonyms and MeSH terms) were combined using appropriate operators. A more limited search was conducted in the Cochrane Library via CENTRAL and CDSR.

Searches in bibliographic databases sought to identify literature published over an appropriately long period of time (inception to the date of search), though relevant literature on the particular treatment combinations is likely to be recent. Manual searches of abstracts from conference proceedings of major conferences covered the period from 2016 to July 2018.

5.1.2 State the eligibility criteria used in the study selection and comment on whether they were appropriate.

Identified studies were assessed against predetermined inclusions and exclusions criteria. These are given in Table 18 (reproducing CS Appendices Table 18).

Table 18: Eligibility criteria for the economic evaluations SLR (reproducing CS Appendices Table 18)

PICOS	Inclusion criteria	Exclusion criteria
Population	• Adult patients (≥18 years)* • Human • Established 1st Line CLL (CLL or B-CLL b-cell CLL or SLL) • With or without del(17p) or *TP53* mutation • ± including patients who are suitable and unsuitable for FCR/BR	• Patients without established 1st line CLL • Paediatric patients (<18 years) • Animal studies • In vitro studies • Patients with aggressive Non-Hodgkin's lymphoma (Richter's transformation or pro-lymphocytic leukaemia)
Intervention	• No restrictions applied	• N/A
Comparator	• No restrictions applied	• N/A

Outcomes	• Total costs • Quality-adjusted life years • ICERs/whether cost effective at some ICER threshold. • Cost per life year gained • Cost per progression free year	• Any outcome not specified under inclusion criteria
Study Design	• Economic Evaluations, such as • Cost utility analysis • Cost effectiveness analysis • Cost minimization analysis	• Economic evaluations not reporting outcomes of interest. • Study designs not specified under inclusion criteria
Publication Type	• Full text articles	• Review articles** • Notes • Erratum • Comments • Editorials • Letters
Language	• Publications in English^	• Publications in any language other than English

*Studies were excluded if the average age of the population is lower than 18. The inclusion of individual patients younger than 18 years of age in an otherwise adult population did not make the article ineligible for inclusion.
**Economic evaluations published in peer-reviewed journals or conference abstract proceedings will be limited to English publications. Evidence from HTA reports will not be restricted to English as it is expected to be published in national languages of the respective HTA agencies.
^Reviews and network meta-analyses will be checked for bibliographic references ONLY and will not be extracted.
Abbreviations: BR: bendamustine in combinations with rituximab; B-CLL: B-Cell Chronic Lymphocytic Leukemia; CLL: Chronic Lymphocytic Leukemia; FCR: Fludarabine, cyclophosphamide and rituximab; ICER: incremental cost-effectiveness ratio. **Source:** AbbVie Data on File (Previously untreated CLL economic SLR report) [10]

As anticipated, certain selection criteria (such as those related to population, comparators, publication type and language) were shared between the clinical effectiveness and cost-effectiveness SLRs. No particular concerns are raised by the ERG in relation to these criteria, though of note is the restriction to studies focusing on adults aged ≥18 years and the exclusion of studies published in languages other than English. The former restriction is in line with the population of participants in CLL14 but is narrower than the NICE scope (which does not specify an age limit), while the latter is a common practice grounded on practical reasons.

Within the cost-effectiveness SLRs, separate sets of inclusion and exclusion criteria were used for selecting literature on HRQoL and health care resource use and costs. While criteria related to population, intervention, comparator and language were identical to those used in identifying relevant economic evaluations (presented in Table 18), some criteria were appropriately different and tailored to capture evidence specific to HRQoL and resource use (e.g. criteria related to outcomes, study design and publication type) (Table 19 and Table 20 below).

Table 19. Eligibility criteria for the health-related quality of life studies (partially reproducing CS Appendices, Table 28)

PICOS	Inclusion criteria	Exclusion criteria
Outcomes	• Disutility and utility measures which comply with one of the following: • Health State Utility values elicited using direct methods: time trade-off and standard gamble • Preference-Based methods: (e.g. EQ-5D, HUI3, SF-6D, aqol, QWB) • Visual analogue scale • Oncology-specific HRQOL tools (e.g.: FACT-Leu; MRC/EORTC QLQ-Leu)	• Any outcome not specified under inclusion criteria
Study Design	• Clinical trials • Observational studies	• Phase I clinical trials • Individual case reports • Systematic Reviews* • Non-systematic reviews* • Genetic/biochemical studies
Publication Type	• Full-text articles • Conference abstracts	• Review articles • Notes • Erratum • Comments • Editorials • Letters

*Reviews and network meta-analyses will be checked for bibliographic references ONLY and will not be extracted. Abbreviations: aqol: assessment of quality of life; B-CLL: b-cell chronic lymphocytic leukaemia; CLL: chronic lymphocytic leukaemia; EQ-5D: EuroQol 5-dimension; FACT-Leu: functional assessment of cancer therapy – leukaemia; HRQoL: Health related quality of life; HUI3: Health utility index 3; MRC/EORTC QLQ-Leu: Medical research council/ European organisation for research and treatment of cancer quality-of-life questionnaire; N/A: not applicable; QWB: Quality of wellbeing; SLL: small cell lymphocytic leukaemia; SF-6D: Short-form 6 dimension; Source: AbbVie Data on File (Previously untreated CLL economic SLR report)

Table 20. Eligibility criteria for the healthcare cost and resource use studies (partially reproducing CS Appendices, Table 32)

PICOS	Inclusion criteria	Exclusion criteria
Outcomes	• Outpatient and inpatient healthcare resource utilisation • Direct costs of inpatient and outpatient services • Indirect costs	• Any outcome not specified under inclusion criteria

	• Costs of adverse events	
Study Design	• Economic evaluations • Patient chart reviews • Patient and disease registry studies • Claims data analyses	• Clinical Trials (Phase I/ II/ III/ IV) • Studies not reporting outcomes of interest
Publication Type	• Full-text articles	• Review articles** • Notes • Erratum • Comments • Editorials • Letters
Language	• Publications in English	• Publications in any language other than English
**Reviews and network meta-analyses will be checked for bibliographic references ONLY and will not be extracted. Abbreviations: B-CLL: b-cell chronic lymphocytic leukaemia; BR: bendamustine and rituximab; CLL: chronic lymphocytic leukaemia; Fludarabine, cyclophosphamide and rituximab, N/A: not applicable; SLL: small cell lymphocytic leukaemia. Source: AbbVie Data on File (Previously untreated CLL economic SLR report)		

Overall, the selection criteria employed are deemed suitable and appropriate for the purposes of the undertaken reviews.

5.1.3 What studies were included in the cost effectiveness review and what were excluded?

The SLRs carried out by the company identified 43 economic evaluations, 20 studies providing information on HRQoL and 16 studies giving estimates of healthcare resource use and costs. Only a small number of these studies was used in the submitted economic analysis. In relation to identified economic evaluations, the company stated that none of the identified studies pertained to the decision problem of interest in this submission and therefore none were directly relevant for decision-making. The ERG concurs that the identified studies do not address the exact decision problem that this technology assessment is concerned with and agree that a *de novo* economic analysis was required.

Information on resource use, costs and HRQoL sourced from the available literature was used in the form of inputs in various components of the economic model, including calculations of costs and estimation of quality-adjusted life years. As anticipated, some key information, including parameter values and assumptions, were also drawn from completed NICE technology

appraisals in CLL. The suitability and appropriateness of using specific pieces of information in respective parts of the economic analysis is critiqued in Section 5.2.

5.1.4 What does the review conclude from the data available?

The SLRs presented in the CS identified a number of studies meeting the inclusion criteria, though these studies do not directly address the decision problem concerning this appraisal. The ERG agrees that a *de novo* economic evaluation tailored to the requirements of the specific final scope and decision problem was necessary. While there is no paucity of information on a number of aspects, such as costs and health state utility values (HSUVs), in the public domain (including peer-reviewed publications and previous NICE technology appraisals (TAs)), much of this information has not been produced with the specific decision problem in mind and its applicability to the submitted economic analysis needs to be judged on a case-by-case basis. Nevertheless, the ERG believes that, using existing published evidence (e.g. in peer-reviewed studies and previous NICE TAs) can serve as useful input in the submitted economic model.

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

As part of their submission to NICE, the company made available a description of their economic analysis and a decision model developed and presented in Microsoft Excel®. This is referred to as the 'original CS model'. Updated models, based on the original model but featuring amendments, were also submitted alongside other evidence in the following instances:

- As part of the company's responses to the ERG's request for clarifications on the original CS. This is referred to as the 'original CS clarification responses model'.
- As part of the updated analysis, submitted as an addendum (August 2019 data cut). This is referred to as the 'CS addendum model'.
- As part of the company's responses to the ERG's request for clarifications on the CS addendum. This is referred to as the 'CS addendum clarification responses models'.

A summary and critique of the submitted economic evidence is presented below.

5.2.1 NICE reference case checklist

The NICE Reference Case checklist is given in Table 21 below.

Table 21: NICE Reference Case checklist

Element of health technology assessment	NICE Reference Case	Does the submission adhere adequately to the Reference Case?
Defining the decision problem	The scope developed by NICE	Yes (also see discussion about differences in Section 3 above)
Comparator(s)	As listed in the scope developed by NICE	Yes (also see discussion in Section 3.3 above)
Perspective on outcomes	All direct health effects, whether for patients or, when relevant, carers	Yes
Perspective on costs	NHS and Personal Social Services	Yes
Type of economic evaluation	Cost–utility analysis with fully incremental analysis	Yes
Time horizon	Long enough to reflect all important differences in costs or outcomes between the technologies being compared	Yes
Synthesis of evidence on health effects	Based on systematic review	Yes. A systematic review was conducted. Key information is drawn from data collected in the CLL14 trial. Further information and model parameters have been obtained from the existing literature and available NICE Single Technology Appraisals.
Measuring and valuing health effects	Health effects should be expressed in QALYs. The EQ-5D is the preferred measure of health-related quality of life in adults.	Yes
Source of data for measurement of health-related quality of life	Reported directly by patients and/or carers	Partially (the utility value used for 'progression-free, off IV treatment' was elicited through vignettes)
Source of preference data for valuation of changes in health-related quality of life	Representative sample of the UK population	Yes
Equity considerations	An additional QALY has the same weight regardless of the other characteristics of the individuals receiving the health benefit	Yes

Evidence on resource use and costs	Costs should relate to NHS and Personal Social Services resources and should be valued using the prices relevant to the NHS and Personal Social Services	Yes
Discounting	The same annual rate for both costs and health effects (currently 3.5%)	Yes
Abbreviations: NHS: National Health Service; NICE: National Institute for Health and Care Excellence; QALY: quality-adjusted life year.		

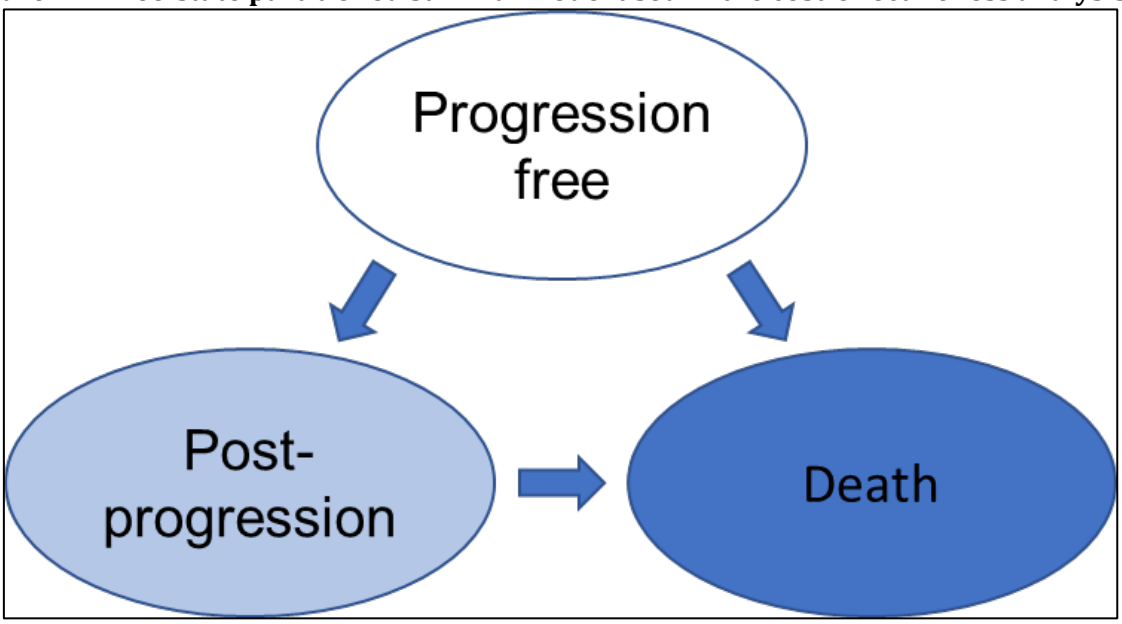
5.2.2 Model structure

The economic model submitted as part of the CS follows a partitioned survival approach and comprises three health states:

- Progression-free survival (PFS, also referred to as pre-progression state), which is populated by CLL patients who are alive and have not progressed.
- Post-progression survival (PPS, also referred to as post-progression state), which includes CLL patients who are alive but have progressive disease, and
- Death, which is the final, absorbing state populated by deceased CLL patients.

The model has a maximum time horizon of 30 years (in the base-case analysis) and is evaluated over a series of cycles, each lasting 28 days. The model's cycle length matches the dosing schedule (length of treatment cycles) for VenG and its comparators. A half-cycle correction has been used in the calculations. The company's representation of the model structure is given in Figure 1 (reproducing Figure 18 in the CS) below.

Figure 1: Three-state partitioned survival model used in the cost-effectiveness analysis



The model structure depicted above was used for both comparisons presented in the CS, namely (i) VenG vs. GClb in patients without del(17p)/TP53 mutation for whom FCR/BR treatment is unsuitable, and (ii) VenG vs ibrutinib in patients with del(17p)/TP53 mutation. Briefly, patients enter the model in the PFS state, where they receive one of the first-line treatments. Patients remain in the PFS state until they die or experience disease progression, upon which event they transition to the Death or the PPS state, respectively. Patients in the PPS state may remain in the state or die, in which case they reach the absorbing state - Death. The proportion of the modelled cohort within each of the three health states at each point in time is guided by extrapolated PFS and OS curves. Of note is the fact that initiation of subsequent treatment is informed by TTNT curves, rather than assumed to take place instantaneously upon progression to PPS. Death due to causes other than CLL (i.e. background mortality) was guided by age-adjusted and sex-adjusted mortality risk values drawn from UK life tables published by the Office for National Statistics [33].

On the whole, the ERG believes that the type and structure of the submitted model is appropriate for the purposes of the condition investigated and adequate for the decision problem considered in this appraisal. The pathway employed in the model is in line with expectations around the clinical progression of the disease, while the structure of the model is generally suitable for capturing and quantifying key costs and health outcomes associated with the compared treatments. A more complex structure (for example, a structure employing sub-states that further distinguish between patients on-treatment or off-treatment in various health

states) may have resulted in a more accurate representation of patients' experience, however the ERG considers that the submitted, parsimonious model is adequate and appropriate for this appraisal.

5.2.3 Population

As discussed in Section 3.1 above, the CS, including the submitted models, relates to a narrower population than that specified in the NICE decision problem. It focuses on VenG in two populations:

- Patients with previously untreated CLL, without del(17p)/TP53 mutation, with known comorbidities that make them unsuitable for treatment with FCR/BR.
- Patients with previously untreated CLL, with del(17p)/TP53 mutation.

The CS populations are similar to those defined as relevant subgroups for consideration in the NICE final scope for this appraisal (See Section 3.5 above) nonetheless, in the CS they constitute the key population groups of interest. Mutation status (i.e. del(17p) and/or TP53 mutation being present) have been combined into a single sub-population, on the premise that these mutations are known to share similar prognostic information. Different algorithms were used to categorise patients to mutation status for use in the CSR and the cost-effectiveness model (CEM).

In response to ERG requests for clarity, the company stated that CLL14 trial participants were assigned to the two mutation status subgroups according to the following algorithm:

- If del(17p) is abnormal (determined by central lab), variable = 1
- If del(17p) is normal (determined by central lab), variable = 0
- If del(17p) is missing & TP53 is mutated, variable = 1
- If del(17p) is missing & TP53 is unmutated, variable = 0
- Else if both are missing = NA

While the ERG accepts that it is reasonable to combine mutation status involving del(17p)/TP53 mutation into a single variable, the ERG sought to understand the company's rationale for using different approaches for mutation status categorisation in the CSR and CEM analyses and the impact that this may have on the results. Clarifications were sought from the company in

relation to subgroup numbers in the CSR and CEM analysis presented in the original CS and the CS addendum, and the discrepancies between them (see Table 22).

In their response, in addition to alluding to errors in the CSR that have been addressed in a corrigendum, the company stated that differences in approaches to combining del(17p)/TP53 mutation status between the CSR analysis and the CEM analysis are due to the latter being derived from a programming method that was used for combining del(17p)/TP53 mutation status in the CEM analysis for NICE TA561.[34] It was further clarified that *"the CEM analysis algorithm prioritises the del (17p) status, whereas the CSR analysis algorithm considers the del(17p) status and TP53 status individually."* (CS addendum clarification response B6).

Table 22: Population numbers utilised in the CSR and CEM analyses

	CSR Analysis **(original CS, Table 32)**	CSR Analysis **(CS addendum, Table 16)**	CEM Analysis **(original CS, Table 32)**	CEM Analysis **(CS addendum, Table 16)**
Non-del(17p) / *TP53* mutation	386	368	387	391
Del(17p) / *TP53* mutation	46	49	31	31
Undefined	0	15	14	10
Total	432	432	432	432
Abbreviations: CEM: cost-effectiveness model; CSR: clinical study report.				

To evaluate the extent to which the use of a different algorithm for CEM impacts on the calculated results, the ERG asked the company to provide a new version of the economic model where cost-effectiveness results were calculated according to time-to-event (TTE) inputs accruing from the CSR (rather than the CEM) categorisation. Using the CSR categorisation in the model resulted in a small difference in the cost-effectiveness results for VenG against its comparators GClb and Ibrutinib in the non-del(17p) and the del (17p) model populations, respectively. The ERG accepted the CEM mutation status categorisation as an appropriate basis for economic modelling.

5.2.4 Interventions and comparators

The comparisons addressed in the company's economic submission and the corresponding populations are described below in Table 23 (reproducing Table 1 in the CS).

Table 23: Sub-populations considered in this submission

Population	Comparison	Rationale
Subpopulation 1: Patients with previously untreated CLL, without del(17p)/*TP53* mutation, with known comorbidities that make them unsuitable for treatment with FCR/BR	VenG vs GClb	• This subpopulation best reflects the cohort of the pivotal trial, CLL14 • The subpopulation is consistent with NHS clinical practice; clinical experts treating patients with CLL in the UK NHS have confirmed that VenG would not be used in

		patients suitable for fludarabine- or bendamustine-based therapies
Subpopulation 2: Patients with previously untreated CLL, with del(17p)/*TP53* mutation	VenG vs ibrutinib monotherapy	• This subpopulation is also reflected in the pivotal trial, CLL14, where 10.6% of patients has del(17p)/*TP53* mutation • There is a high unmet need for this poor-prognostic subpopulation
Abbreviations: CLL: chronic lymphocytic leukaemia; GClb: chlorambucil with obinutuzumab; NHS: National Health Service; VenG: venetoclax with obinutuzumab.		

In the economic model, venetoclax is administered as an oral tablet over a fixed treatment duration of 12 cycles. The treatment is delivered with an initial dose escalation schedule (from 20mg to 400mg daily over cycles 1 and 2, followed by 400mg daily over cycles 3-12). Obinutuzumab is administered as an intravenous infusion, over a fixed treatment duration of 6 cycles, in line with the administration schedule in CLL14.

In the company's analysis, chlorambucil (Clb) is considered to be administered orally over 12 cycles, as per the treatment schedule in the control arm of CLL14. This schedule, however, is not aligned with Clb use in UK clinical practice, where the drug is typically administered over six cycles. The company acknowledged this discrepancy and argued that the overall dose, which is likely to have a larger impact on efficacy than the number of cycles, is broadly similar to the overall dose administered in UK clinical practice. According to the company, experts at an AbbVie-organised HTA advisory board opined that the difference in the number of cycles should not be a concern because 12 cycles of GClb as used in the control arm of the CLL14 trial would most likely lead to better results than six cycles, making the modelled comparison more favourable to GClb and conservative for VenG. No further evidence was offered to substantiate this opinion.

Advice sought from the ERG's clinical expert confirmed that the overall dose is likely to have a larger impact on efficacy than the number of cycles, thus the regimens are comparable in terms of efficacy. The ERG's clinical expert considered the assumption of equivalence between the six and 12-cycle Clb regiments to be reasonable and opined that most CLL experts would advocate 12 months of chlorambucil-based therapy. However, our expert confirmed that, in UK clinical practice, Clb is typically offered over six cycles. In light of this, and to identify the impact of this shorter treatment schedule, the ERG requested an additional analysis where Clb is administered over six cycles. In their response, the company pointed to variability in the UK practice in relation to Clb dose and number of cycles and provided alternative versions of the model based

on a six cycle treatment schedule and alternative doses and points of treatment delivery within each cycle (0.5mg, 0.10 mg, on days 1 and 7 or 1 and 15). It is noted that changing the number of cycles from 12 to six in the model required the doubtful assumption that the efficacy benefit of a 12-cycle Clb treatment (rather than a possibly reduced 6-cycle efficacy benefit) as per the CLL14 trial outcomes is maintained.

The change to a six-cycle schedule had a very small impact in the calculated cost-effectiveness results. In consultation with the NICE Technical Team, it was agreed that the original analysis, based on the 12-cycle treatment schedule, provides appropriate and informative findings, thus the main results are provided on the basis of the 12-cycle treatment schedule.

5.2.5 Perspective, time horizon and discounting

The analysis is presented from an NHS and Personal Social Services perspective, in line with the NICE reference case. Patients are modelled over a 30 year time horizon, which for a typical cohort of CLL patients effectively constitutes a lifetime horizon. In the base-case, costs and benefits were discounted at an annual rate of 3.5%.

5.2.6 Treatment effectiveness and extrapolation

The company use four time-to-event outcomes from CLL14 as inputs into the economic model: PFS, OS, TTNT, and time on treatment (ToT). Extrapolations were assessed through a combination of information criteria, Akaike and Bayesian information criteria (AIC and BIC), assessment of visual fit and assessment of the clinical plausibility of long-term predictions. Each extrapolation was subject to constraints in case the hazard rate of the extrapolation was too optimistic.

For each defined population (i.e. patients without del(17p)/TP53 mutation and patients with del(17p)/TP53 mutation), we present a summary of the company's implementation of each input based on their CS addendum, accompanied by the ERG's critique and recommendations.

5.2.6.1 Non-del(17p)/TP53 mutation population: progression-free survival

Proportionality

The company assessed and rejected the assumption of proportional hazards between the two arms of CLL14 for PFS. This was done through an assessment of Schoenfeld residuals and a

formal test of proportionality. The test for proportionality did not lead to a rejection of the hypothesis that proportional hazards held (p=[redacted]), though the threshold for significance was not clearly stated. However, the Schoenfeld residual plot had a clear curvilinear trend suggesting that the proportionality assumption did not hold. An examination of the log-cumulative hazard plot showed that the curves crossed at roughly 6.5 months and then gradually diverge also suggesting that the hazard rates were not proportional. Whilst the population that proportionality was evaluated on included both patients with and without del(17p)/TP53 mutation, the ERG have no reason to believe that the proportionality assumption between treatments would be different for these two groups, and agree that proportionality across the follow-up period does not hold.

Despite the likely violation of proportional hazards, using dependent models does have other advantages. Given the immaturity of the data, it can be beneficial to use data from both arms to ensure hazard rate behaviour is consistent and plausible for both arms. The company utilised this approach in the appraisal for venetoclax-rituximab for second-line CLL,[34] where they assumed proportionality between PFS and OS and between both treatments simultaneously. By making these assumptions, one can reduce the uncertainty around each extrapolation, through the borrowing of information.

Constraints

The PFS extrapolations were subject to the following constraints:

- there could not be more patients in the progression-free health state than there were alive.
- the hazard rate of disease progression could not fall below the hazard rate of background mortality.

The ERG accepts the rationale behind the first constraint, which is routinely implemented in partitioned survival models such as this. However, the reasoning for the second constraint is less clear. The ERG believes that it may not be true that the PFS event rate should always be higher than background mortality. The ERG accepts that the overall mortality rate of the CLL14 trial should never drop below that of background mortality. However, it is the healthiest patients who will remain in the progression-free health-state population, and the ERG finds it plausible that they may have a progression/mortality rate below that of background mortality. This constraint applied by the company may unnecessarily increase the PFS event rate, potentially introducing bias into the cost-effectiveness analysis.

Extrapolations

The company compared standard parametric extrapolations fitted separately to each arm of the PFS data: exponential, Weibull, log-logistic, log-normal, generalised gamma and Gompertz curves. The effect of the del(17p)/TP53 mutation was modelled through the inclusion of a covariate in the parametric model. This assumed proportionality of either the hazard rate or the failure time, depending on the parametric curve being fitted. This meant that the effect of the del(17p)/TP53 mutation was modelled separately for both arms. This is of concern to the ERG, as it reduces the information contributing to each effect estimate suggesting each will be surrounded by large uncertainty, though this uncertainty was not made clear by the company. There is also no evidence either in support or against the assumption of an interactive effect between the deletion/mutation and treatment.

Figure 2, taken from the CS, demonstrates that there are clear differences between the PFS extrapolations for both arms, but particularly among the VenG extrapolations. Note that the extrapolations in this figure have not been subject to the constraints mentioned above.

Figure 2: PFS extrapolations from CLL14 for patients without del(17p)/TP53 mutation, unconstrained

The company presents the AIC and BIC statistics for the parametric curves, shown here in Table 24, and long-term predictions for the extrapolations (Table 25), which have had the constraints applied, explaining the inconsistency with Figure 2.

The constraints implemented by the company come into effect in both arms of the trial. For VenG, the background mortality rate is used instead of the PFS extrapolation from 10.9 years in the economic model, and for GClb from 16.9 years.

Table 24: AIC and BIC for PFS models fitted independently to arms of CLL14 trial

Distribution	AIC		BIC	
	VenG	GClb	VenG	GClb
Exponential	[redacted]	[redacted]	[redacted]	[redacted]
Weibull	[redacted]	[redacted]	[redacted]	[redacted]
Gompertz	[redacted]	[redacted]	[redacted]	[redacted]
Log-logistic	[redacted]	[redacted]	[redacted]	[redacted]
Log-normal	[redacted]	[redacted]	[redacted]	[redacted]
Generalised gamma	[redacted]	[redacted]	[redacted]	[redacted]

Table 25: PFS predictions from parametric models fitted to CLL14 trial and benchmarks

Distribution	VenG				GClb			
	3 year	5 year	10 year	20 year	3 year	5 year	10 year	20 year
Exponential	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Weibull	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Gompertz	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Log-logistic	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Log-normal	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Generalised gamma	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
CLL11 (95% CI)	-	-	-	-	42%	25% (19-31)	-	-
ERIC study	-	-	-	-	42%	-	-	-
CLL14	81.9%	-	-	-	49.5%	-	-	-
ERG Clinical Expert	75%	50%	20%	5%	40%	25%	0%	0%
2 knot hazard spline	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Both AIC and BIC suggest that the exponential curve is the best fit to the VenG data, but that the log-logistic curve is the best fit to the GClb arm.

The company states that their statistical experts advised that statistical fit should not be relied upon when data are immature. The ERG agrees and are reluctant to allow AIC and BIC to influence the choice of curve given the immaturity of the data.

The company rejects the exponential curve as selecting this for both arms would result in an assumption of proportional hazards which has already been concluded as false. The ERG also agrees with this consideration, also noting the exponential extrapolation for GClb is too optimistic.

The company then compares the extrapolation to the observed data from CLL11[35], since both trials had a GClb treatment arm. The log-logistic and the Weibull models provide the closest predictions to the observed 5 year PFS from CLL11, however slightly overestimate and underestimate respectively. The company reports selecting the log-logistic for their base case analysis through consideration of the goodness of fit combined with an examination of the plotted hazard functions. However, the company does not elaborate on exactly how the hazard functions influenced their decision making process. The company concludes by reporting that clinical experts validated the log-logistic curve as the closest to clinical practice for GClb patients.

The ERG note that the independent extrapolations considered by the company all overestimate the three year PFS rate in both arms, and considerably overestimate the three year PFS rates for the comparable GClb groups from CLL11 and the ERIC study[26] (see section 4.8), and the predictions of the ERG's clinical expert.

The ERG has some concerns with the company's justification in selecting the log-logistic curve. Firstly, the data immaturity means extrapolation with any curve is unlikely to accurately capture the true survival profile for patients in either arm.

Secondly, when the ERG considers the hazard function of the log-logistic curve, without constraints applied, the nature of the log-logistic curve is to model a decreasing hazard rate beyond the tail of the Kaplan-Meier (KM) plot for both arms. The ERG finds this implausible, given that OS events for progression-free patients are included in this measure. This view appears to be supported by the company in their original submission. When assessing the proportionality assumption for TTNT using the previous data cut, the company states that they

expected the assumption to be rejected due to the "close correlation" between PFS and TTNT. They later reject the log-logistic as a candidate model for TTNT due to the decreasing hazards over time since it is "clinically implausible". It is unclear why the company was willing to select the log-logistic curve for PFS but not for TTNT, despite the correlation between the outcomes.

The fact that the hazard rates for the extrapolations of both arms fall below background mortality reinforce the ERG's view that the extrapolations are unsuitable.

The ERG also compared the mean PFS time from the company's base case analysis, and contrasted it to that from the appraisal of GClb for the same indication (TA343). Both the company's and the ERG's base case from the initial review of TA343 estimated the undiscounted mean time in the PFS health state to be 2.83 life years (LYs). In the present appraisal, the log-logistic extrapolation preferred by the company estimates this mean time for PFS to be [redacted] LYs.

The ERG is reluctant to recommend any of the independent PFS extrapolations presented by the company based on the immaturity of the data and the implausibility of the extrapolations, and so we considered alternative extrapolations.

The alternative approaches made available by the company in the economic model included the option to use spline models to extrapolate PFS or to model using KM data followed by a parametric extrapolation in a piecewise approach, however these were not discussed in the CS.

For completeness, the ERG investigated the plausibility of the extrapolations for 60 PFS models, consisting of the combinations of dependent and independent, parametric and spline models, fitted with and without the KM data. The extrapolations that were used after an initial period of KM data were still fitted to the full observed set of data, and not only to the tail data. The ERG found that these models apply the predicted hazard rate to the observed data from 28 months, and were just a small step change from the extrapolations that are performed without prior modelling of the KM data. It is unclear why the company selected the cut-off of 28 months, however this setting could be varied within the economic model.

The ERG found that the spline model with 2 knots fitted on the hazard scale predicted a 5 year PFS rate of [redacted]% for GClb, the closest of all independent models to the observed data from CLL11. The corresponding estimate for mean PFS was [redacted] LYs, an improvement over the

company's base case, closer resembling estimates from the extrapolations of TA343. The 10 year estimate for VenG PFS was also consistent with the prediction of the ERG's clinical expert.

These extrapolations for PFS could still be overoptimistic in terms of their estimates of proportion estimated to be progression-free and mean PFS time for both arms, but are the most plausible from the data in its current state of maturity.

5.2.6.2 **Non-del(17p)/TP53 mutation population: Time-to-next treatment**

Time-to-next treatment is used in combination with OS to estimate a pseudo-health-state that the ERG will refer to as "time on next treatment" (TONT). This is similar to the commonly utilised health-state in partitioned survival models of post-progression survival, in that its population size is calculated as the difference between the TTNT curve and the OS curve. However, TONT is only used to calculate the number of patients on second line treatment, and the associated costs of this treatment. There is no utility value attached to this group of patients, as this is dictated by their progression status. The TTNT is a very influential parameter in this appraisal, due to the high costs of later line therapies. As with PFS, the immaturity of the data raises concerns over the accuracy and reliability of any extrapolation.

The TTNT proportionality assessment and extrapolation were performed in a similar manner to PFS. The company rejected the proportionality assumption from examination of the observed cumulative hazard plots and residual plots. The ERG agrees with this interpretation of the evidence, though note there is a stronger argument for proportionality for TTNT than PFS, and maintain the view that the assumption may be helpful when extrapolating given the immaturity of the data. Recall that the company treated death events as TTNT events, which may confound the extrapolations.

The company considered both parametric and spline models fitted independently to both arms of CLL14.

<u>Constraints</u>

TTNT extrapolations were constrained by the following rules:

- The number of patients who had not begun their next treatment could not exceed the number of patients alive.

- The hazard rate of beginning next treatment could not fall below the hazard rate of background mortality.

Extrapolations

The company examined the AIC and BIC for all models. The exponential model was the best fit for the VenG arm. For GClb, the log-logistic is the best fit when considering both AIC and BIC. As before, the immaturity of the data leaves the ERG unwilling to rely heavily, if at all on AIC and BIC for the selection of a curve for extrapolation.

The company then compares the extrapolations to the CLL11 trial, which reported at 5 years that 49% of GClb patients had not experienced a TTNT event. In their first submission, the company rules out the log-logistic curve due to its "clinically implausible" decreasing hazard rate over time, and selected from the generalised gamma, Weibull and spline models the model with the best statistical goodness-of-fit, leaving them with the Weibull. However, in their addendum using the extended follow-up, the company prefers the log-logistic model, contradicting their earlier justification.

The company now states in the CS Addendum that the log-logistic curve aligns with the clinical expectation for patients in remission, however the ERG remains unconvinced of this justification.

The ERG is not supportive of the use of the log-logistic curve or any of the independent extrapolations for TTNT considered by the company, as no curve provides 5 year estimates which are comparable to what was observed in CLL11 (see Table 26).

Firstly, the data are immature meaning the estimated parameter values will be associated with large uncertainty and the fitted models are unlikely to capture the true behaviour of TTNT. Secondly, the constraint of the hazard rate for TTNT to not fall below background mortality is necessary and comes into effect at just before 6 years for the VenG log-logistic extrapolation and just before 14 years for the GClb log-logistic extrapolation. This implies the rate of patients beginning a second treatment or dying before beginning a second treatment is equivalent to the rate of background mortality, casting further doubt on the reliability of the extrapolations.

In search of an alternative, the ERG investigated 60 dependent and independent models with and without using KM data before the extrapolations to obtain more plausible estimates. Only two models produced estimates of 5 year TTNT that were within ± 5% of the 49% observed in CLL11. Both of these models were independently fitted probit spline models with 2 knots. Whilst the ERG interprets the resulting estimates for GClb to be plausible for these two probit spline models, the predictions for VenG were still constrained by background mortality within the first 7 years of the economic model.

The ERG also requested that the company implement extrapolations of TTNT where OS events were instead censored rather than counting as events. However, these extrapolations were no more plausible than the original ones provided by the company.

The ERG considered an alternative approach, and recreated the patient level data of CLL14 data for TTNT and PFS provided by the company within the economic model using the ipdfc command in Stata 16.[36] The ERG hypothesized that due to the similarities between TTNT and PFS, that the assumption of proportionality might hold between these outcomes, and fitted a stratified Cox proportional hazard model, stratifying by arm of the trial. A check of the cumulative hazard plots suggested that the assumption did hold (Figure 3). This produced a hazard ratio of [redacted] that the ERG applied to the hazard rates from the ERG preferred PFS curve in the economic model. The resulting 5 year prediction was within the 95% confidence interval of the estimate from CLL11, and the constraint to background mortality only came into effect beyond 20 years for both arms, where its influence was much smaller. Hence, the ERG prefers to use this approach to extrapolate TTNT over the parametric models provided by the company.

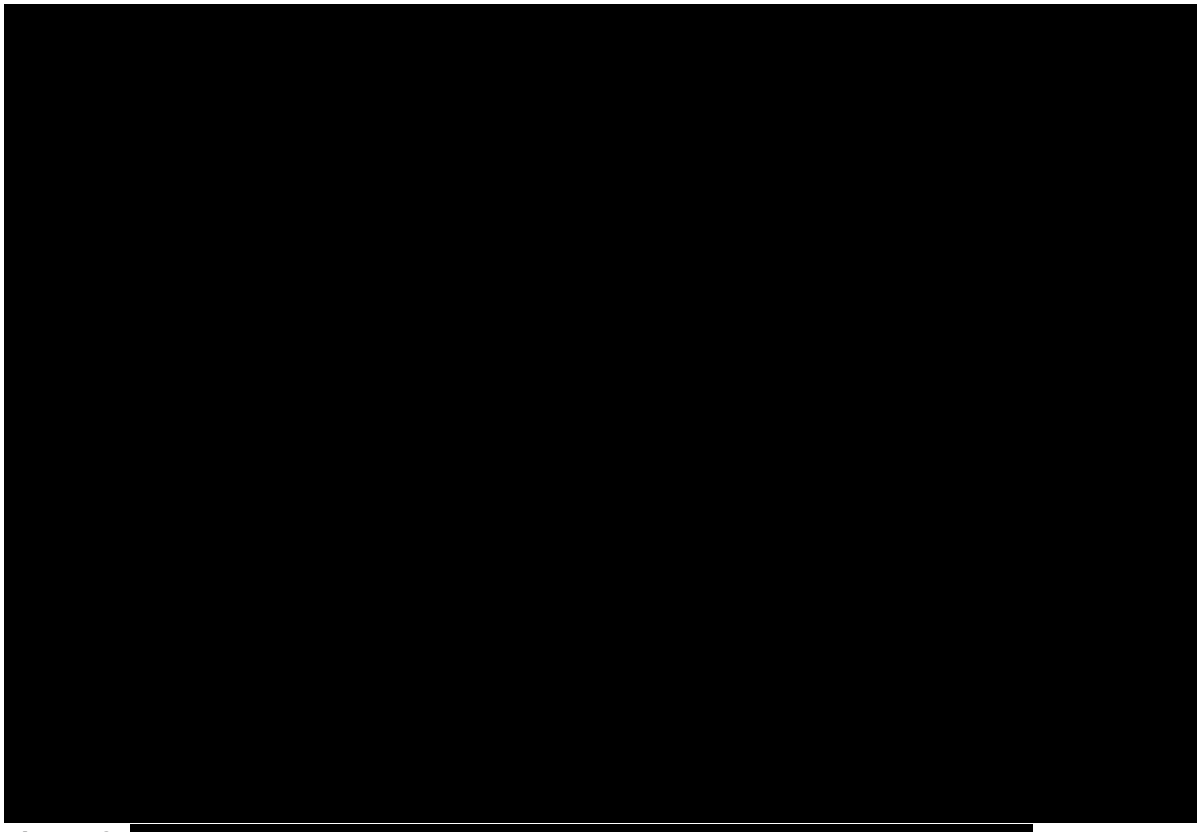


Figure 3:

Table 26: Predictions of TTNT from CLL14 data for non-del(17p)/TP 53 mutation population

Distribution	VenG			GClb		
	5 year	10 year	20 year	5 year	10 year	20 year
Exponential						
Weibull						
Gompertz						
Log-logistic						
Log-normal						
Generalised gamma						
CLL11 Data	-	-	-		-	-
ERG hazard ratio on PFS extrapolation						

5.2.6.3 **Non-del(17p)/TP53 mutation population: overall survival**

The company performed model selection and extrapolation for OS in an identical way to PFS and TTNT. The only constraint applied was that the hazard rate of OS could not fall below the rate of background mortality. Due to the immaturity of the data, the company assumes that the

OS for the VenG arm is equivalent to that of the GClb arm, despite including a covariate for the effect of VenG in the model. This assumption of equivalent for OS also seemingly disregards the strong benefit of VenG that was observed and modelled for PFS and TTNT. Examination of the limited follow-up suggested the data were consistent with both the assumptions of proportionality and equivalence, but it is unclear whether these assumptions are truly appropriate for either the observed period or the extrapolated period, given the immaturity of the data.

The company details their model selection, considering AIC, BIC and the hazard behaviour of the various contender models. However, this is largely irrelevant as can be seen when comparing the long-term predictions from the models in Table 27. Most of the models appear to produce very similar extrapolations, despite immature data usually being associated with large uncertainty. This similarity is due to the effect of the background mortality constraint, and not because the parametric curves necessarily agree. For the exponential model chosen by the company in their base-case, the background mortality rate comes into effect from 4.87 years. All of the models have background mortality coming into effect at a similar time, suggesting none of the extrapolations would be considered plausible without this constraint.

The company fitted models either independently to both arms, or simultaneously to both arms with a covariate for treatment effect. For their base-case, the company used an exponential model and fitted a dependent model, with a parameter for first line treatment. They then used the predicted curve for the GClb arm to represent OS for both arms.

The ERG is unclear why the company did not refit the model removing the parameter, adjusting for the arm of the CLL14, when assuming equivalence.

In their original submission, the company cites other studies with longer-term follow-up. The study with the highest absolute survival of 53% at 10 years was Shvidel et al[37] however, this study included patients who were eligible for FCR and who are likely to be a healthier population, though this effect could be negated by developments in later lines of therapy that would affect patients in CLL14.

The ERG's clinical advisor provided comment on the fact that the company's modelling of OS suggested that the OS from CLL14 was almost identical to that of the general population

meaning there is no additional risk of death from CLL, and stated this to be "untrue". The ERG also compared the company's predicted OS to the OS observed in the 5-year follow-up of CLL11. Whilst there likely are differences between the studies in terms of the later lines of therapy available, the ERG is unconvinced that this can explain the considerable difference between the company's predictions and the data observed in CLL11.

Hence the ERG is reluctant to select any of the curves considered by the company as they all provide implausible extrapolations and rely heavily on the constraint of background mortality.

Table 27: Overall Survival predictions from dependent parametric models for non-del(17p)/TP 53 mutation population

Distribution	GClb (also used for VenG)			
	3 year	5 year	10 year	20 year
Exponential	[redacted]	[redacted]	[redacted]	[redacted]
Weibull	[redacted]	[redacted]	[redacted]	[redacted]
Gompertz	[redacted]	[redacted]	[redacted]	[redacted]
Log-logistic	[redacted]	[redacted]	[redacted]	[redacted]
Log-normal	[redacted]	[redacted]	[redacted]	[redacted]
Generalised gamma	[redacted]	[redacted]	[redacted]	[redacted]
Background mortality	[redacted]	[redacted]	[redacted]	[redacted]
CLL11 GClb	[redacted]	[redacted]	[redacted]	[redacted]
ERIC Study GClb	[redacted]	[redacted]	[redacted]	[redacted]
ERG Clinical Expert		[redacted]	[redacted]	[redacted]
ERG OS using ERIC hazard rate from 3 years	[redacted]	[redacted]	[redacted]	[redacted]

In pursuit of a plausible extrapolation, the ERG considered the 60 curves incorporated by the company into the economic model, as performed for the previous time-to-event outcomes. However, none predicted 5-year survival of below 80%, and were therefore not considered consistent with the CLL11 data by the ERG.

Next, the ERG digitised KM plots and recreated patient level OS data from the CLL11 trial, and obtained patient level data from the investigators of the ERIC study. Exponential models were fitted to the data from both trials to investigate their hazard rates, however these models were

then inconsistent with the observed data from CLL14. In order to obtain a model that was consistent with both CLL14 and the external studies, the ERG considered using piecewise modelling, where the hazard rate from CLL14 was modelled for the first three years of the economic model, followed by the hazard rate from the ERIC study. The ERIC study was preferred over CLL11 since the data are more recent, and the later lines of therapy likely to be more consistent with what would be received by patients in the UK moving forward. However, the hazard rates were similar from both studies. The ERG's clinical expert also stated that the OS data from the ERIC study is "very believable and representative" of patients under this indication, given that it is real world data.

The ERG maintained the assumption of equal OS between the two arms, because although it is plausible that VenG could offer some benefit, there is no clinical evidence to support this or provide any quantification of this benefit. The ERG's clinical expert also commented that there is no evidence yet on how effective salvage therapies are after first line VenG, whereas ibrutinib is demonstrated to be effective following GClb, so it is unclear whether this assumption could be considered conservative. Under the ERG's preferred OS assumption, the constraint of background mortality comes into effect at 14 years.

5.2.6.4 **Non-del(17p)/TP53 mutation population: Time on treatment**

The time spent on first-line treatment outcome was not extrapolated, as the trial follow-up exceeded two years and all patients had discontinued first-line treatment within the observed period. The company modelled the observed data, however it was subject to the constraint that the proportion of patients stopping their first treatment could not exceed the proportion that had begun their second treatment, according to the modelling of TTNT. The company also adjusted ToT without providing clear justification. Instead of using the data as observed from CLL14, they restricted ToT such that no patients in the economic model could exceed the licensed duration of 12 cycles. However, following consultation with our clinical expert, the ERG concludes that it is likely these patients still appearing on treatment had experienced a pause of treatment, and to extend the time on first line treatment would effectively double count treatment costs for these patients. Hence, the ERG is satisfied with the company's modelling of time on treatment.

5.2.6.5 **Del(17p)/TP53 mutation population: VenG**

For the time-to-event extrapolations, the company estimated the efficacy of VenG in the del(17p)/TP53 mutation population by including the relevant covariate into the model for the

non-del(17p)/TP53 mutation population as discussed above. The rationale for doing this is that the size of this subgroup in CLL14 was too small to extrapolate from and this approach allowed the company to borrow information from the wider trial population. The company appeared to include patients who received GClb in their analysis, despite these patients being irrelevant and potentially misleading for the eventual comparison to ibrutinib. This approach makes the assumption of proportionality between the two subgroups of patients, difficult in addition to other assumptions of proportionality or equivalence made between treatment arms.

For PFS and TTNT the company used the models in the same manner as in the above sections. For OS, the company fitted the same dependent model, but this time included the covariate for the differing effect of VenG relative to GClb when predicting OS for VenG patients, in addition to including the deletion covariate. Hence, the OS for VenG patients with del(17p)/TP53 mutation was not assumed to be equivalent to the same subgroup receiving GClb. Recall, that for the population without del(17p)/TP53 mutation, the company assumed that OS for VenG patients would be equal to that of GClb patients, despite including a treatment effect parameter for VenG in their model. The rationale for the inconsistency in this assumption is unclear as it is not discussed by the company.

In response to the ERG's clarification request following the company's addendum, the company assessed proportionality between the subgroups using plots of the Schoenfeld residuals, ignoring potential treatment effects. There was no clear evidence of violation of proportionality for PFS or OS, though the small sample size of the del(17p)/TP53 mutation group makes it difficult for the ERG to be confident that proportionality is a reasonable assumption to make. For TTNT, the company only assessed proportionality for the previous data cut (August 2018), however it appeared to support the assumption of proportional hazards.

The company compared the visual fit of the extrapolations to the KM curves for VenG patients from CLL14. The ERG interpreted these and concluded that the fit to all three outcomes could be considered reasonable (Figure 4, Figure 5 and Figure 6).

In their response to the ERG clarification questions referring to the original submission, the company provided detailed model output for PFS and OS which suggested that the hazard rates of these outcomes for the del(17p)/TP53 mutation subgroup were significantly different to the hazard rates for the rest of the CLL14 population.

Figure 4: PFS from CLL14 for del(17p)/TP53 mutation population - company base case

Figure 5: TTNT from CLL14 for del(17p)/TP53 mutation population - company base case

Figure 6: OS from CLL14 for del(17p)/TP53 mutation population - company base case

5.2.6.6 **Del(17p)/TP53 mutation population: ibrutinib**

Given the lack of a direct comparison between ibrutinib and VenG in the del(17p)/TP53 mutation population, the company applied the hazard ratios estimated from their naïve indirect comparisons onto the extrapolations for VenG. A comparison of the log-cumulative hazard plots comparing data from Mato et al[38] suggested that the proportional hazards assumption for PFS and OS was violated, though this could be influenced by the small sample sizes. If violated, this would leave all comparisons presented by the company to be unreliable as they all assume proportionality between ibrutinib and VenG for PFS and OS.

As before PFS and OS hazard rates were constrained by background mortality, and there could not be more patients in the progression free health state than alive. The constraint with background mortality was not necessary as extrapolated mortality rates remained above background mortality for the duration of the economic model.

Time on treatment for ibrutinib was modelled to be equivalent to PFS, which is consistent with how ibrutinib is currently administered.

TTNT is not explicitly modelled for ibrutinib. This means the way the company modelled the time on next treatment for ibrutinib was inconsistent compared to VenG and GClb. For ibrutinib, the company counted only new incidences of either progressive disease events or death events within each cycle as contributors to the time-on-next-treatment. This means that across the time horizon of the model, each patient received later line therapy for a single cycle. The rationale for not considering the possibility of remaining longer on next treatment is not provided by the company, and remains unclear to the ERG.

The company's analysis was also found to have further flaws, which became apparent when investigating the Markov trace plots, which track the health-state of the population through the time horizon of the economic model.

In Figure 7 showing the Markov trace for VenG, it is apparent that there is [redacted]. Similarly, for ibrutinib, Figure 8, patients spend [redacted]. It is possible that the lack of a post-progression health state is what led the company to their unusual approach for modelling time-on-next-treatment for ibrutinib patients. The ERG is concerned with the implausibility of these outcomes from the company's analysis.

Figure 7: Markov trace for VenG in del(17p)/TP53 mutation population

Figure 8: Markov trace for ibrutinib in del(17p)/TP53 mutation population

The ERG attempted to investigate whether either of the PFS and OS extrapolation could be considered more reliable than the other, as both contribute to the construction of the post-progression health state (Table 28). A comparison of the company's predictions to the ERG's clinical advisor suggests that both the PFS and OS extrapolations may be too optimistic, and the ERG considered alternative curves.

Table 28: Comparison of PFS and OS estimates between company's and ERG's base case time-to-event outcomes for 17p deletion/TP53 mutation population.

PFS					
	Company			**ERG**	
	VenG (independent log-logistic extrapolation from CLL14)	Ibrutinib (Hazard ratio of 0.66 applied to VenG extrapolation)	ERG Clinical Expert Prediction (same for both treatments)	VenG (1 knot hazard spline, independent)	Ibrutinib (Hazard ratio of 0.66 applied to VenG extrapolation)
5 year	[redacted]	[redacted]	10%	[redacted]	[redacted]
10 year	[redacted]	[redacted]	0%	[redacted]	[redacted]
20 year	[redacted]	[redacted]	0%	[redacted]	[redacted]

OS					
	VenG (exponential dependent extrapolation from CLL14)	Ibrutinib (Hazard ratio of 0.84 applied to VenG extrapolation)	ERG clinical Advisor OS (same for both treatments)	VenG (1 knot hazard spline, dependent)	Ibrutinib (Hazard ratio of 0.84 applied to VenG extrapolation)
5 year	[redacted]	[redacted]	40%	[redacted]	[redacted]
10 year	[redacted]	[redacted]	10%	[redacted]	[redacted]
20 year	[redacted]	[redacted]	0%	[redacted]	[redacted]

The ERG found that the 1-knot hazard spline model produced estimates that were closer to the predictions of their clinical expert. These also predicted that patients in both arms would have a more plausible duration in the post-progression period. (Figure 9 and Figure 10). Whilst the extrapolations used by the ERG appear more plausible than those presented by the company, the lack of meaningful data informing both the VenG extrapolation and the indirect comparison to ibrutinib mean the ERG is hesitant to recommend these assumptions for consideration for decision making.

The ERG also preferred to use the 1 knot hazard dependent spline model for VenG TTNT as this predicted that later lines of therapy would be taken for [redacted] years, rather than [redacted] years as under the company's assumptions. The ERG was unable to change this duration for ibrutinib patients due to the company's approach to modelling, however any improvement would only increase the costs of later line therapy associated with ibrutinib which were underestimated, suggesting any estimate of cost-effectiveness of VenG may be conservative. In both ERG and company base-cases, the average time on later lines of therapy for first-line ibrutinib patients was [redacted] years.

Figure 9: Markov trace plot for VenG del(17p)/TP53 mutation patients under ERG assumptions.

Figure 10. Markov trace plot for ibrutinib del(17p)/TP53 mutation patients under ERG assumptions.

Summary

A summary of the ERG's preferred assumptions for the modelling of time-to-event outcomes in the non-del(17p)/TP53 mutation and del(17p)/TP53 mutation populations can be found in Table 29 and Table 30, respectively.

Table 29: ERG's preferred assumptions in relation to time-to-event outcomes in non-del(17p)/TP53 mutation population.

Outcome	Company Base Case for non-del(17p)/TP53 mutation	ERG Base Case for non-del(17p)/TP53 mutation
PFS	Independent log-logistic extrapolation of CLL14 data	Independent 2-knot hazard spline extrapolation of CLL14 data
TTNT	Independent log-logistic extrapolation of CLL14 data	Hazard ratio between TTNT and PFS calculated from recreated CLL14 data applied to ERG PFS extrapolation.
OS	Used OS exponential extrapolation of GClb data from CLL14 for both arms	Used exponential model fitted to IPD from ERIC study to extrapolate beyond 3 years from CLL14 data.
TOT	Data from CLL14 capped at 12 months	Same as company

Table 30: ERG's preferred assumptions in relation to time-to-event outcomes in del(17p)/TP53 mutation population.

Outcome	Company Base Case for del(17p)/TP53 mutation	ERG Base Case for del(17p)/TP53 mutation
PFS	Independent log-logistic extrapolation of CLL14 data and hazard ratio for ibrutinib	Independent 1-knot hazard spline extrapolation of CLL14 data and hazard ratio for ibrutinib
TTNT	Independent log-logistic extrapolation of CLL14 data	Dependent 1 knot hazard spline extrapolation of CLL14 data
OS	Used OS exponential extrapolation of VenG data from CLL14 and hazard ratio for ibrutinib	Dependent 1-knot hazard spline extrapolation of CLL14 data and hazard ratio for ibrutinib
TOT	Data from CLL14 for VenG and equal to PFS for ibrutinib	Same as company

5.2.7 Health related quality of life

5.2.7.1 Health state utility values

Estimates of HRQoL included in the economic model were drawn from two main sources: the available literature and the CLL14 trial. Estimates from the literature were used in the company's main analyses, whereas estimates from CLL14 were used in scenario analyses.

Health status descriptions, key component for constructing preference-based health-related quality of life (utility) indices, were collected as part of the CLL14 trial using the EuroQol EQ-5D-3L instrument. Two models, one that included time as a covariate and one that did not, were used to estimate utility values from the latest available CLL14 data (August 2019 data cut-off). In the CS addendum clarification response B1, values are given for the populations with and

without del(17p)/TP53 mutation; however, these values relate only to the PFS state and were not treatment-arm specific. Utility values based on CLL14 data reported in the CS addendum clarification response B1 are given in Table 31.

Table 31. Estimated PFS utility values from CLL14 (August 2019 data cut-off)

	With del(17p)/TP53	Without del(17p)/TP53
Model 1*	[redacted]	[redacted]
Model 2 (with time) **	[redacted]	[redacted]
*Derived from $U_{it}= \alpha+\beta_1 txarm_i+\beta_2 age_i+ \beta_3 sex_i+\varepsilon_{it}$		
** Derived from $U_{it} = \alpha + \beta_1 txarm_i + \beta_2 age_i + \beta_3 sex_i + \beta_4 cycle_t + \varepsilon_{it}$		

PFS: progression-free survival.

Table 32 (Table 6 in the CS addendum clarification response B2) shows EQ-5D scores derived from data collected in CLL14 (August 2019 data cut) at different states (baseline, pre and post progression) by mutation status categorisation and treatment arm. It must be noted that values for GClb patients in the del(17p)/TP53 mutation population (3rd column in the Table) is irrelevant to this decision problem, as GClb is not an investigated comparator in this sub-population. The value set ('tariff') used to translate status descriptions to preference-based indices is not stated implicitly, but it is assumed that this was the time-trade-off based UK specific value set for EQ-5D-3L.[39]

Table 32: EQ-5D utility values from CLL14 trial (August 2019 data cut-off)

EQ-5D scores in CLL14	With del(17p)/*TP53*		Without del(17p)/*TP53*	
	VenG	GClb	VenG	GlCb
Baseline (Cycle 1 Day 1)				
Number of responses/patients that responded	[redacted]	[redacted]	[redacted]	[redacted]
Mean value (Standard deviation)	[redacted]	[redacted]	[redacted]	[redacted]
Progression-free				
Number of observations	[redacted]	[redacted]	[redacted]	[redacted]
Number of eligible patients to respond	[redacted]	[redacted]	[redacted]	[redacted]
Mean value (Standard error)	[redacted]	[redacted]	[redacted]	[redacted]
Post-progression				
Number of observations	[redacted]	[redacted]	[redacted]	[redacted]
Number of patients that responded	[redacted]	[redacted]	[redacted]	[redacted]

Mean value (Standard error)	[redacted]	[redacted]	[redacted]	[redacted]

Abbreviations: EQ-5D: European Quality of Life 5 Dimensions; GClb: chlorambucil with obinutuzumab; VenG: venetoclax with obinutuzumab.

While a large proportion of CLL14 trial participants contributed HRQoL data at baseline and at points in time before progression, only a small number of patients recorded in the available dataset (August 2019 data cut) had progressed, thus HRQoL estimates for the post-progression state are subject to considerable uncertainty. The company noted that differences between arms are not statistically significant across populations, as p-values of the treatment arm coefficient were consistently above 0.05 in all regression analyses of the CLL14 EQ-5D-3L data. The ERG agrees that, in light of the reported results, it is sensible that the base-case analysis is based on non-treatment-specific utility values.

Based on advice from clinical and health economic experts at an AbbVie-organised advisory board, the company argued that utility values from the CLL14 trial were unfeasibly high, as they exceeded the age and gender-matched values for the general population (70-year old - female 0.77, male 0.79). Thus, rather than using CLL14 data, a decision was made to use PFS and PPS health state utilities from published sources. The ERG considers the rationale for not using the unexpectedly high utility values from CLL14 to be reasonable and in line with arguments previously accepted in a previous CLL-related appraisal.[32]

A SLR was carried out (conducted in December 2018 and updated in July 2019) to identify relevant HSUVs in patients with previously untreated CLL (discussed in Section 5.1 above). According to the CS and the company's subsequent answers to a request for further clarifications, only three[40-42] of the identified publications in the HRQoL-specific SLR reported EQ-5D values. All three publications report on the same study, Connect CLL. HRQoL data in this study were collected from US individuals and the exact EQ-5D value set used to derive utility indices is not reported. The ERG considers that it is unlikely that a UK values set was used to derive utilities and agrees that these studies are very unlikely to be in line with the NICE reference case.

Given the above, the company sought an alternative source of utility values for the previously untreated CLL population and opted to use values from NICE TA343[32] as these have previously been accepted as plausible by NICE. While the original CS (and the original CS model) did not use separate utility values for different treatment status within health states, the company amended their approach and provided a new analysis that was based on using separate utility

values within the PFS health state to take into account patients' HRQoL while they are on and off treatment. The utility values used in the company's base-case analysis are presented in Table 33 below.

Table 33. New utility values suggested in the company's response to ERG's clarification questions

Progression stage	Utility value	Source	Rationale for change/use
Pre-progression IV	0.670	TA343 for PFS under IV treatment	VenG and GClb include IV treatment. This is applied for the fixed treatment duration of 12 months in the PFS state.
Pre-progression off treatment	0.820	TA343 for PFS after initial treatment is completed (0.82)	VenG and GClb should not be taking into account IV disutility for the complete time on PFS health state. A value higher than that of pre-progression oral treatment (0.71) treatment but lower than that of perfect health is a more suitable option.
Pre-progression oral treatment	0.710	TA343 for PFS under oral treatment	A utility value reflective of oral treatment should be applied for the Ibrutinib arm PFS state.
Post-progression	0.600	TA343: weighted average of the following utilities (progression after first-line treatment, PFS ± second-line treatment, relapsed line of treatment)	Used as base case and aligned with what has been accepted in previous NICE CLL appraisals.[32]

These values, which have been previously used in the economic model submitted as part of TA343 (obinutuzumab in combination with chlorambucil for previously-untreated chronic lymphocytic leukaemia)[32], were obtained from a utility elicitation study carried out by Roche®. The study aimed to derive societal preferences for QoL associated with CLL and involved eliciting utility scores from 100 members of the public for nine health states descriptions (vignettes).[32] It must be noted that the ERG which undertook TA343 considered the data from the Roche® study to be of low quality, as HRQoL was not elicited directly from patients using a generic questionnaire, such as the EQ-5D. In particular, while the ERG accepted that in the absence of a better quality of life data, Roche's study should inform the utility values, they raised a concern about the utility value of 0.82 used for progression-free patients when off treatment, as this was higher than the age-adjusted values for members of the UK general public (in the particular case, 0.76). The ERG suggested that a utility of 0.76 should be seen as an upper bound.

The ERG believes that including utility values for sub-states is appropriate; however, we question the value chosen for PFS utility off treatment, which has been criticised and retracted in TA343[32] and contradicts the rationale that CLL patients are unlikely to have better quality of life than non-CLL patients of the same age and gender in the general population. Thus, the ERG's preferred approach is to cap the value for the 'PFS, off treatment' sub-state to the gender-weighted, age-specific value for members of the UK general public. To estimate this value, the ERG followed a simple approach, based on the formula for calculating utility (EQ-5D) values in the general population published by Ara and Brazier.[43] The same formula has been used in the CS for the calculation of age adjusted general population values.

$$General\ Population\ index = 0.9508566 + 0.0212126 \times male - 0.0002587 \times age - 0.0000332 * age^2$$

The 'PFS, off treatment' value applies after patients on VenG and GClb therapies have completed their first line treatment. Given that in these therapies first line treatment is provided over a fixed duration of approximately 12 months, the ERG estimated the utility value for a 72-year-old (i.e. starting age of 71 years plus 12 months of treatment) member of the general population, calculated according to the male/female population split in the company's submission. This value was estimated to be 0.7703. It is likely that the true value for the utility in PFS after treatment will be lower than this value, though in the absence of specific data, the ERG has adopted this pragmatic approach, which is in line with the approach taken in TA343 appraisal.[32]

A number of sensitivity and scenario analyses using alternative utility values were presented in the original CS (CS Section B.3.8.). Additional analyses were provided in response to the ERG's requests. These included scenario analyses using the treatment arm specific progression-free and post-progression utility values calculated from the CLL14 trial (original CS clarification response B3) and alternative utility values for PFS and PPS retrieved from the literature (original CS clarification response B5). In general, the results showed little change in the magnitude of the incremental effects in both sup-populations of interest and no change in direction compared to the base-case analysis results. However, when using the per arm 'progression free' and 'post-progression' utilities calculated from the CLL14 trial (August 2019 data cut), the difference in quality-adjusted life years (QALYs) between VenG and GClb in the non-del(17p)/TP53 mutation population is reduced substantially as compared to the base-case results.

5.2.7.2 **Disutility due to adverse events**

The expected impact of adverse events (AEs) on patients' quality of life was accounted for by combining utility AE specific utility decrements (disutilities) with duration estimates reflecting the period of time over which each AE is anticipated to affect a patient's HRQoL. Multiplying the disutility value per adverse event by the duration of the AEs gave a QALY decrement which was applied to the first cycle in the model. Disutility values were sourced from previous NICE technology appraisals and existing literature; these are presented in Table 34 (reproducing original CS Table 47). The ERG considers the evidence used to be robust and the sources of this evidence appropriate. Exploratory analyses carried out by the ERG showed disutility values and the calculated QALY decrements to have a very small impact on incremental QALYs and overall cost-effectiveness results. While second line treatment disutilities due to AEs were not included, these omissions are expected to have an inconsequential impact on the difference in QALYs between the compared treatments.

Table 34: Disutility values and QALY decrements due to adverse events

AE	Disutility (positive)	SE	Duration (days)	SE	QALY decrement	Reference
Asthenia	0.115	0.012	35.33	3.54	0.011	NICE appraisal TA306;[44] Lloyd et al. 2006;[45] PIX301 trial
Diarrhoea	0.080	0.005	3.50	0.35	0.001	NICE appraisal TA216;[46] Beusterien 2010;[47] NICE appraisal TA344[48]
Dyspnoea	0.103	0.010	12.70	1.27	0.004	NICE appraisal TA306; [44] Lloyd et al. 2006; [45] PIX301 trial
Febrile neutropenia	0.150	0.015	3.50	0.35	0.001	Lloyd et al. 2006;[45] NICE appraisal TA344[48]
Infusion related reaction	0.200	0.020	3.50	0.35	0.002	NICE appraisal TA344[48]
Leukopenia	0.090	0.009	14.00	1.40	0.003	Assumed to be the same as neutropenia; PIX301 trial
Neutropenia	0.090	0.002	3.50	0.35	0.001	Nafees et al. 2008;[49] NICE appraisal TA344[48]
Pneumonia	0.195	0.004	18.21	1.82	0.010	Tolley et al. 2013;[50] NICE appraisal TA359[51]
Sepsis	0.195	0.004	7.00	0.70	0.004	Tolley et al. 2013; [50] UK NHS Adboard
Thrombo-cytopenia	0.108	0.011	23.20	2.32	0.007	Tolley et al. 2013; [50] NICE appraisal TA359[51]

Abbreviations: AE: adverse event; QALY: quality-adjusted life year; SE: standard error.

5.2.8 Resources and costs

The following key categories of resource use and costs have been included in the company's analysis: (i) intervention and comparators' costs (including treatment acquisition and administration costs, routine care costs, tumour lysis syndrome monitoring costs and subsequent treatment costs), (ii) costs related to adverse events, and (iii) terminal care costs.

5.2.8.1 Intervention and comparators' costs

Unit costs of drugs comprising VenG and its comparators were sourced from the British National Formulary (BNF) online database.[52] An overview of the treatment regimens modelled in the analysis, as well as the drug acquisition cost (per pack size and per mg) are reproduced in Table 35 and Table 36, respectively (reproducing original CS Tables 50 and 49).

Table 35. Treatment regimens for VenG and comparators.

Regimen	Drug	Admin	Dosing schedule	Cost per cycle	Trial name (Reference)
VenG	Venetoclax	Oral	Venetoclax: • 20 mg daily during Cycle 1, Days 22–28 • 50 mg daily during Cycle 2, Days 1–7 • 100 mg daily during Cycle 2, Days 8–14 • 200 mg daily during Cycle 2, Days 15–21 • 400 mg daily during Cycle 2, Days 22–28 and on Days 1–28 for all subsequent cycles until the end of Cycle 12	Cycle 1, Days 22–28: £59.87 Cycle 2, Days 1–7: £149.67 Cycle 2, Days 8–14: £299.34 Cycle 2, Days 15-21: £598.68 Cycle 2, Days 22–28: £1,197.37 Cycle 3–12: £4,789.47	CLL14 [7]
	Obinutuzumab	IV	• 100 mg on Day 1 and 900 mg on Day 2 (or 1000 mg on Day 1) • 1000 mg at Cycle 1, Day 8 and Day 15 • 1000 mg at Day 1 for all subsequent cycles until the end of Cycle 6	£9,936 for Cycle 1 £3,312 for Cycle 2–6	CLL14 [7]
GClb	Obinutuzumab	IV	• 100 mg on Day 1 and 900 mg on Day 2 (or 1000 mg on	£9,936 for Cycle 1	CLL14 [7]

			Day 1) • 1000 mg at Cycle 1, Day 8 and Day 15 • 1000 mg at Day 1 for all subsequent cycles until the end of Cycle 6	£3,312 for Cycle 2–6	
	Chlorambucil	Oral	0.5 mg/kg at Day 1 and Day 15 for Cycles 1–12	Assuming a weight of 76: £64.79	CLL14 [7]
Ibr	Ibrutinib	Oral	420 mg daily continuously (until evidence of progressive disease or no longer tolerated by the patient)	£4,292	RESONATE [53]
Abbreviations: GClb: chlorambucil with obinutuzumab; Ibr: ibrutinib; IV: intravenous; VenG: venetoclax with obinutuzumab.					

Table 36. Drug costs for venetoclax and comparators

Drug	Dose per tablet or vial	Units per package	Cost per package	Price per mg	Source
Venetoclax Tablet, mg	10 mg	14	£59.87	£0.43	BNF [52]: Venclyxto (AbbVie Ltd)
	50 mg	7	£149.67	£0.43	
	100 mg	7	£299.34	£0.43	
	100 mg	14	£598.68	£0.43	
	100 mg	112	£4,789.47	£0.43	
Obinutuzumab, IV, mg/ml	1000mg	1	£3,312.00	£3.31	BNF[52]: Gazyvaro 1000mg/40ml concentrate for solution for infusion vials (Roche Products Ltd)
Chlorambucil, Tablet, mg	2 mg	25	£42.87	£0.86	BNF[52]: Chlorambucil 2mg tablets (Alliance Healthcare (Distribution) Ltd)
Ibrutinib, Tablet	140 mg	90	£4,599.00	£0.37	BNF[52]: Imbruvica 140mg capsules (Janssen-Cilag Ltd)
	140 mg	120	£6,132.00	£0.37	
Abbreviations: BNF: British National Formulary; IV: intravenous.					

The original CS stated that there is a simple discount patient access scheme (PAS) for venetoclax which entails providing a discount of [redacted] on the list price for venetoclax.

5.2.8.2 **Administration costs**

Administration costs were included in the model for the intravenously-delivered treatments obinutuzumab and rituximab (subsequent treatment) (Table 51 in the original CS, reproduced as Table 37 below). Cost calculations for treatment administration accounted for the cost of pharmacist time for dispensing the IV drugs.[54] Alternative delivery methods (standard IV, rapid IV and subcutaneous administration) were also considered. Some assumptions were employed to enable calculations, including: (i) that the cost of a rapid infusion would be similar to a simple chemotherapy delivery included in the NHS reference costs, and (ii) that rituximab containing treatment (VenR) uses a 30:70 ratio between standard and rapid IV infusions. The latter was justified on the basis of a survey exploring administration policies that was conducted in 20 UK trusts. The administration cost assigned to obinutuzumab was that of a standard IV infusion.

Table 37. Drug administration costs.

Drug	Cost	Currency code	Description
IV standard	£298.53 (= £289.33 + £9.20)	SB15Z	IV administration cost from NHS Reference Costs 2017-18; Total HRGs, SB15Z: deliver subsequent elements of a chemotherapy cycle. This is supplemented by the cost of pharmacist time for dispensing the IV drugs (£9.20).
Rituximab (IV Rapid)	£238.19 (= £228.99 + £9.20)	SB12Z	IV administration cost from NHS Reference Costs 2017-18; Total HRGs, SB12Z: deliver Simple Parenteral Chemotherapy at First Attendance. This is supplemented by the cost of pharmacist time for dispensing the IV drugs (£9.20).
Abbreviations: IV: intravenous; HRG: Healthcare Resource Group; NHS: National Health Service.			

The ERG considers the methods and assumptions employed in calculating administration costs to be reasonable. Sensitivity analyses using alternative values for administration cost inputs (costs, split between rapid and standard infusion) demonstrated a very small impact on total incremental costs and overall cost-effectiveness results.

5.2.8.3 **Routine care and monitoring costs**

A range of health care resources associated with the pre-progression and post-progression states were included in the cost calculations. These included scans, blood tests, transfusions and consultations and were informed by discussion with clinicians at an AbbVie-organised advisory board. While the clinical expert supporting the ERG considered the type and frequency of care comprising the annual resource use reasonable, the ERG identified some discrepancies in categories of routine care included and annual frequency of use between this and a previous

submission (TA561[34]). Values for both submissions are given in Table 38 below. Checks carried out by the ERG using the categories and values specified in TA561[34] led to a small increase in the difference in total costs between treatments in favour of VenG.

Table 38. Pre- and post-progression annual resource use frequency

Resource/procedure	TA 561 [34]		CS, Table 52	
	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	2
Chest x-ray	0	2	0	-
Bone marrow exam	0	1	0	-
Haematologist visit	2	6	4	4
Inpatient non-surgical medical stays	0	4	0	3
Nurse home visit	-	-	-	-
Full blood transfusion	0	11	0	1
Biochemistry test: renal - Urea and electrolytes test	-	-	3	2
Biochemistry test: liver function test	-	-	3	2
Immunoglobulins Blood Test	-	-	3	2

Abbreviations: CT: computerised tomography; LDH: lactate dehydrogenase.

National reference costs available for the most recent year (2017/18)[55] were used to inform the routine care and monitoring costs. These are given in Table 39 (reproducing original CS Table 53).

Table 39: Routine care and monitoring costs used in the model

Routine care and monitoring costs	Value	HRG codes from reference costs 2017/18[555552525355]
Full blood count	£2.51	DAPS05- Haematology
LDH	£1.11	DAPS04 - Clinical biochemistry
Haematologist visit	£159.65	Outpatient Attendances Data: 303- Clinical haematology
Inpatient non-surgical/medical visit	£572.78	National schedule of reference costs 2017/18: Weighted average of day case, Chronic Lymphocytic Leukaemia, including Related Disorders, SA32A (£396), SA32B (£428), SA32C (£379) and SA32D (£449) = £432.93

		PSSRU 2018: Medical consultant hour (£108) + qualification costs (£31.846) = £139.846
Full blood transfusion	£187.97	Outpatient Procedures- 303, Clinical Haematology, single plasma exchange or other intravenous blood transfusion, 19 years and over (SA44A)
CT Scan	£92.81	Weighted average of RD20A (£88) and RD21A (£106)[29]
Biochemistry test: renal - Urea and electrolytes test	£1.11	DAPS04 – Clinical biochemistry
Biochemistry test: liver function test	£1.11	DAPS04 - Clinical biochemistry
Immunoglobulins Blood Test	£2.51	DAPS05- Haematology (assumed to be equal to full blood count)
Abbreviations: CT: computerised tomography; HRG, Healthcare Resource Group; LDH, Lactate Dehydrogenase; PSSRU, Personal Social Services Research Unit.		

5.2.8.4 Treatment-specific monitoring costs – Tumour lysis syndrome

CLL patients are at increased risk of tumour lysis syndrome (TLS), a condition that occurs when a large number of cancer cells die within a short period. TLS is most commonly observed in patients with hematologic malignancies and, although the risk of TLS in CLL is deemed to be small[56], developing the condition can have a significant impact on health and economic outcomes. Thus, the expected cost of TLS prophylaxis was calculated and included in the model. Calculations were based on an algorithm that categorised patients according to risk of developing TLS based on data from the treated CLL14 population (August 2019 data cut-off).

In brief, patients were divided into those at lower and greater risk based on tumour mass and absolute lymphocyte count (ALC) (i.e. lower risk: lymph node with a diameter ≤5 cm and ALC $<25 \times 10^9$/L; greater risk included all other patients). This resulted in [redacted] of patients on VenG and [redacted] patients on GClb being included in the lower risk group, and 86.57% of patients in the VenG arm and 87.96% of patients in the GClb arm being part of the greater risk group. The greater risk group was subdivided into two groups according to creatinine clearance.

The TLS risk group distribution and the cost by risk of tumour burden can be seen in Table 40 and Table 41, reproducing Tables 32 and 33 in CS addendum.

Table 40. TLS risk distribution for VenG and GClb treatment arms

Treatment	Lower Risk (node diameter ≤5 cm and ALC <25 x 10^9)	Greater Risk (node diameter >5 cm or ALC >25 x 10^9)	
		Creatinine clearance > 80 mL/min	Creatinine clearance ≤ 80 mL/min
VenG	[redacted]	[redacted]	[redacted]
GClb	[redacted]	[redacted]	[redacted]

Abbreviations: ALC: absolute lymphocyte count; GClb: chlorambucil with obinutuzumab; VenG: venetoclax with obinutuzumab.

Table 41: TLS cost split by tumour burden in each treatment arm.

Treatment	Low tumour burden	Greater Risk (CrCl >80)	Greater Risk (CrCl >80)	Total cost used in model
VenG	£1,411	£1,708	£2,230	£1,784
GClb	£1,411	£1,489	£2,242	£1,629

Abbreviations: CrCl: creatinine clearance; GClb: chlorambucil with obinutuzumab; VenG: venetoclax with obinutuzumab.

According to calculations based on the TLS risk classification and the prophylaxis algorithm, the cost of TLS prophylaxis applied in the first cycle of the model to the VenG arm and GClb is £1,784 and £1,629 respectively. The company stated that the cost is lower in the GClb arm as there are fewer high-risk patients in the GClb arm compared to the VenG arm. This appears to favour GClb and is therefore a conservative assumption for VenG. The TLS costs of the venetoclax regimens were halved, doubled and removed in scenario analyses. The ERG believes that the approach and inputs used in the TLS prophylaxis calculations is reasonable and broadly in line with previously submitted evidence for TA561.[34] Checks carried out by the ERG confirmed the company's assertion that these costs have a small impact on overall cost-effectiveness.

5.2.8.5 Subsequent treatment costs

Subsequent treatment costs were calculated according to the type of subsequent treatment mix received, the point in time when subsequent treatment would be initiated and the length of time over which subsequent treatment would be administered.

The company determined the type of treatment mix offered to patients after first line treatment by consulting UK-based clinical experts. Subsequent treatments included in the economic

model, stratified by population (with and without del(17p)/TP53 mutation) and first line treatment (VenG, GClb, ibrutinib) are reproduced in Table 42. Clinical expert advice sought by the ERG confirmed that these treatments are consistent with what is offered in UK clinical practice. The ERG's clinical expert added that venetoclax with rituximab (VenR) is becoming increasingly more popular as a subsequent treatment for patients who had VenG or GClb as first line treatment, as offering VenR instead of ibrutinib means that the option of offering ibrutinib remains available, should it be needed as a further treatment. Our expert indicated that a reasonable expectation would be that, in the future, about 20% of patients would be offered ibrutinib after first line treatment, and 80% would be offered VenR. However, a change in these proportions has a minimal effect on overall incremental results.

Table 42. Overview of base case subsequent treatment mix

Initial treatment	**Subsequent treatment**	
	Non-del(17p)/*TP53* mutation	**del(17p)/*TP53* mutation**
VenG	50% ibrutinib; 50% VenR	100% ibrutinib
GClb	50% ibrutinib; 50% VenR	N/A
Ibrutinib	N/A	100% venetoclax monotherapy
Abbreviations: GClb: chlorambucil with obinutuzumab; VenG: venetoclax with obinutuzumab; VenR: venetoclax with rituximab.		

The median length of time (in months) over which subsequent treatment is received were sourced from recent published literature identified through a systematic review (Table 43).

Table 43: Subsequent treatment durations used in the model

Subsequent treatment	**Median duration, months**	**Source**
VenR	24.4	Kater et al. (2019) [57]
Ibrutinib	39.00	O'Brien et al. (2018)[58]
Venetoclax monotherapy	16.00	Davids et al. (2018) [59]
Abbreviations: VenR: venetoclax with rituximab.		

Together with the treatment mix and split in Table 42 above, these lengths were used to calculate the average treatment acquisition cost per cycle over the subsequent treatment period. The point in time at which eligible patients would receive subsequent treatment was estimated according to the TTNT curves for VenG and GClb after these were adjusted for overall survival from the CLL14 trial.

A different approach was used in estimating subsequent treatment costs for ibrutinib in the del(17p)/TP53 mutation population. There, the proportions of patients who receive subsequent treatment and are still alive was obtained based on the PPS and OS curves. The company justified the use of a different approach by explaining that publicly available patient level data to inform TTNT curves for ibrutinib could not be identified.

5.2.8.6 Adverse reaction unit costs and resource use

An overview of adverse event costs was given in CS, Table 61 (partially reproduced in Table 44 below). The company stated that these were aligned with the accepted costs used in TA561, with minor changes made to the costs for neutropenia, leukopenia, diarrhoea, and sepsis according to clinical feedback at an AbbVie-organised advisory board. The ERG found no further information in the original CS about the reasons for these changes.

Table 44. Adverse event cost overview

Adverse event	Cost	Source
Asthenia	£657.76	TA498: National Schedule of Reference Costs 2017-18, PSSRU 2018[60]
Diarrhoea	£0.34	TA344[48] Woods et al. (2012)[61]
Dyspnoea	£591.49	NHS Reference Costs 2017-18. Total - HRGs, Other Respiratory disorders without interventions (weighted average of DZ19L-DZ19N [£1,132], DZ19M [£725] and DZ19N [£475])[62]
Febrile neutropenia	£6,563.61	NICE TA359: NHS Reference Costs 2012-13; Inflated by four years using the PSSRU HCHS index (£5993.03*1.026*1.019*1.022*1.025).[62]
Infusion related reaction	£432.93	NHS reference costs 2016-2017. Weighted average of day case, Chronic Lymphocytic Leukaemia, including Related Disorders, SA32A (£396), SA32B (£428), SA32C (£379) and SA32D (£449).
Leukopenia	£535.56	Same as neutropenia
Neutropenia	£535.56	Cost of lenograstim for 8 days (median duration of neutropenia in MURANO trial - Seymour et al. 2018)[63]
Pneumonia	£6167.48	NHS Reference Costs 2017-18[55]
Sepsis	£6167.48	Same as pneumonia
Thrombocytopenia	£640.09	NHS Reference Costs 2017-18[55]

Abbreviations: BNF: British National Formulary; HCHS: hospital and community health services; HRG: Healthcare Resource Group; NHS: National Health Service; PSSRU: Personal Social Services Research Unit.

The ERG checked these costs and compared them with AE related costs used in TA561.[34] These are presented in Table 45.

Table 45. AE costs in current appraisal and TA561

Adverse event costs	Current appraisal (CS)	TA561[34]
Asthenia	£657.76	-
Diarrhoea	£0.34	-
Dyspnoea	£591.49	-
Febrile neutropenia	£6,563.61	-
Infusion related reaction	£432.93	-
Leukopenia	£535.56	-
Neutropenia	£535.56	£119.49
Pneumonia	£6,167.48	£6,149.58
Sepsis	£6,167.48	-
Thrombocytopenia	£640.09	£621.34

Between the two submissions, there is a discrepancy in the unit costs for neutropenia, though replacing the value used in the economic model by that accepted in TA561 had a negligible effect on total and incremental costs.

5.2.8.7 Miscellaneous costs

Costs associated with terminal care were calculated and included in the model in the same way as in NICE appraisal TA561.[34] Briefly, terminal care costs were applied to all patients who transition to the death health state as a one-off cost and came from a published study on end-of-life care for solid tumour cancer patients[64] on the basis that the costs of terminal care would be similar between solid tumour and haematology patients. The mean total cost due to terminal care was estimated to be £6,662. No costs related to specific health-states were included in the economic analysis; NHS and Personal Social Services costs accruing over the course of the modelled time horizon are included in the categories above.

5.2.9 Cost-effectiveness results

The company presented base-case results generated from the economic model for the two populations of interest:

- Patients with previously untreated CLL, without del(17p)/TP53 mutation, with known comorbidities that make them unsuitable for treatment with FCR/BR.
- Patients with previously untreated CLL, with del(17p)/TP53 mutation.

Results for two different pricing arrangements were provided: i) list price for VenG vs list price of all comparators (GClb and ibrutinib), and ii) PAS price for venetoclax only (obinutuzumab remains at list price) vs list price of all comparators (GClb and ibrutinib). For brevity, and after

discussion with the NICE Technical Team for this appraisal, it was established that a PAS discount applied to venetoclax only provides uninformative results, the focus in the remainder of this report is on the analyses based on list prices for all treatments. The ERG has produced a confidential addendum reporting cost-effectiveness results calculated based on the venetoclax PAS discount and the confirmed PAS discounts for obinutuzumab and ibrutinib.

In general, the company's preferred base-case analysis suggests that VenG is associated with a greater number of QALYs and lower costs against its comparator in the non-del(17p)/TP53 mutation population, suggesting that VenG is dominant versus GClb. In the del(17p)/TP53 mutation population, VenG resulted in a lower average number of QALYs and lower costs versus ibrutinib. In both cases, VenG resulted in a large, positive net monetary benefit (NMB) at both the £20,000 and £30,000 per additional QALY willingness-to-pay (WTP) thresholds (Table 46).

The company stated that, in the non-del(17p)/TP53 mutation population, cost-effectiveness is largely driven by the superior progression-free survival associated with VenG, and lower subsequent costs following progression for VenG compared to GClb. In the del(17p)/TP53 mutation population, a key driver of the cost-effectiveness results is the much higher treatment costs due to ibrutinib being offered over a non-fixed and typically long period of time (i.e. until patients' progress) in contrast to VenG, which is provided for a fixed number of cycles. Results calculated on the basis of PAS for venetoclax only (not presented here) were of the same magnitude and direction.

Table 46: Company's base case results at list prices

Treatment	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	ICER (£/QALY)	NMB (WTP: £20k per QALY)	NMB (WTP: £30k per QALY)
Non-del(17p)/*TP53* mutation population							
GClb	[redacted]	6.742				[redacted]	[redacted]
VenG	[redacted]	7.799	[redacted]	1.057	[redacted]		
Del(17p)/*TP53* mutation population							
Ibrutinib	[redacted]	4.153				[redacted]	[redacted]
VenG	[redacted]	3.991	[redacted]	-0.163	[redacted]		

Abbreviations: GClb: chlorambucil with obinutuzumab; ICER: incremental cost-effectiveness ratio; NMB: net monetary benefit; QALYs: quality-adjusted life years; VenG: venetoclax with obinutuzumab; WTP: willingness to pay.
*This is a south west quadrant ICER, where the intervention is cost-effective if the ICER is above the willingness-to-pay threshold. Positive NMB values imply that the intervention is a cost-effective use of NHS resources at the given willingness-to-pay threshold.

5.2.10 Sensitivity analyses

Various types of sensitivity analyses, including probabilistic, deterministic and scenario analyses, were undertaken by the company.

5.2.10.1 Probabilistic sensitivity analysis

Probabilistic sensitivity analyses (PSA) were carried out by assigning distributions to a range of uncertain parameters and randomly sampling from these distributions over 1,000 replications. Information about the type of distributions used can be found in the CS original and addendum models. The company stated that, in cases where uncertainty estimates for parameters were unavailable, the analysis employed a variability (i.e. standard error) estimate of 10% of a parameter's mean value. Clarifications were sought by the ERG about the rationale for using the same standard error (typically 10% of a parameter's mean value) uniformly across parameters as diverse as resource use, probability of events, unit costs and utility values (original CS clarification response B7). In addition, the ERG questioned the fact that uncertainty around the OS and PFS hazard ratios for ibrutinib were set to 10% of the mean despite the fact that more accurate uncertainty values for these parameters were available. In response, the company made changes in the uncertainty values (standard errors) assigned to the hazard ratios mentioned above. The company made no changes to standard errors attached to utility values citing lack of information required for the calculation of appropriate values.

Results of the probabilistic sensitivity analyses generated in the CS addendum model for the comparison between i) VenG and GClb in the population without del(17p)/TP53 mutation and ii) VenG and ibrutinib in the population with del(17p)/TP53 mutation are given below (Figure 11 to Figure 14). As before, for all these comparisons, all medications costs, including those for venetoclax, were kept at list prices.

Figure 11: Scatter plot of probabilistic results on the cost-effectiveness plane for non-del(17p)/TP53 population – list prices

Figure 12: Cost-effectiveness acceptability curves for non-del(17p)/TP53 population – list prices

Figure 13: Scatter plot of probabilistic results on the cost-effectiveness plane for del(17p)/TP53 population – list prices

Figure 14: Cost-effectiveness acceptability curves for del(17p)/TP53 population – list prices

5.2.10.2 **Deterministic sensitivity analyses**

The company carried out a number of one-way sensitivity analyses to identify key model drivers and important sources of uncertainty. In each of these analyses, the central estimate of each base-case parameter was replaced by low and high estimates. Tornado plots showing the first ten uncertain parameters whose impact on the incremental cost-effectiveness ratio (ICER) is the greatest can be seen in Figure 15 and Figure 16 for VenG vs GClb and VenG vs ibrutinib, respectively.

For the comparison between VenG and GClb, the one-way sensitivity analysis calculations suggest that the parameter with the greatest impact on the ICER is the PFS utility value following IV treatment (utility for PFS: Post fixed treatment duration (FTD) IV treatment). Whereas, for the comparison between VenG and ibrutinib, the parameter with the greatest impact on the ICER is the PFS utility value at the time of the FTD period (utility for PFS: FTD IV treatment). While, these one-way sensitivity analyses offer indications on the influence of single parameters on the cost-effectiveness results, these should be seen as 'stress tests' where the lower and upper values substituting a parameter may not be realistic. It must also be noted that one-way sensitivity analyses do not account for interrelations between parameters or the fact that more than one of the parameters will be uncertain at the same time.

Figure 15. Tornado plot of deterministic sensitivity analysis: impact on ICER (VenG versus GClb) for non-del(17p)/TP53 mutation population – list prices.

Figure 16. Tornado plot of deterministic sensitivity analysis: impact on ICER (VenG versus ibrutinib) for del(17p)/TP53 mutation – list prices.

5.2.10.3 **Scenario analyses**

Alternative values for various parameters were considered in the company's scenario analyses. Indicatively, scenarios tested included different discount rates, time horizons, exclusion of TLS prophylaxis costs, various utility values from different sources, different survival models based on a range of distributions for VenG and extreme value scenarios related to the specification of PFS and OS curves. A full list of the variables and approaches subjected to scenario analysis are available in the CS, Table 69. Results of scenario analysis were updated to reflect the most recently available data cut (August 2019) and were summarised in CS addendum (Table 40 for the non-del(17p)/TP53 population and in Table 41 for the del(17p)/TP53 population).

Scenario analyses pertaining to the non-del(17p)/TP53 population suggested that VenG is consistently less costly and more effective when compared to GClb. An exception to this was when the time horizon of the analysis was limited to 5 years. Analyses related to the del(17p)/TP53 population showed that VenG resulted in lower costs and lower QALYs, and an overall positive NMB in all but two scenarios (i.e. when equivalent OS and PFS were assumed and time horizon was limited to 5-years. In these cases, VenG was dominant versus ibrutinib).

As discussed above, additional analyses run in response to the ERG's requests included: (i) a scenario where utility values used in the model are treatment-specific values obtained from the CLL14 trial (August 2019 data cut) (Table 47) and (ii) a scenario where Clb is administered over six cycles, which is typically the case in UK clinical practice (Table 48). These analyses were provided in response to ERG queries B4 and B5 in the CS addendum clarifications, respectively.

In both analyses, results are in broad agreement with those of the company's base-case analysis. However, using utility values from CLL14 in the comparison between VenG and GClb (see Scenario 2 in Table 47) led to a significant decrease in incremental QALYs as compared to the base-case results, and a notable reduction in NMB.

Table 47: Scenario analyses undertaken using the utility values from the CLL14 trial (list prices)

Incremental results of VenG vs comparator	Incremental discounted costs	Incremental discounted QALYs	ICER, £/QALY	Net monetary benefit
Scenario 1: with del(17p)/*TP53* mutation VenG: PFS (IV) = [redacted], PPS = [redacted]				
Base case: with del(17p)/*TP53*	[redacted]	−0.163	[redacted]	[redacted]
Scenario 1: vs Ibrutinib	[redacted]	−0.436	[redacted]	[redacted]
Scenario 2: without del(17p)/*TP53* mutation VenG: PFS (IV and post IV*) = [redacted], PPS = [redacted] GClb: PFS (IV and post-IV*) = [redacted], PPS = [redacted]				
Base case: without del(17p)/TP53	[redacted]	1.057	[redacted]	[redacted]
Scenario 2: vs GClb	[redacted]	0.052	[redacted]	[redacted]

*The same utility value is also applied to post-IV since the literature post-IV utility value is less than the utility generated from the CLL14 trial while on IV.
**This is a south west quadrant ICER, where the intervention is cost-effective if the ICER is above the willingness-to-pay threshold. Positive NMB values imply that the intervention is a cost-effective use of NHS resources at the given willingness-to-pay threshold.

Table 48: Scenario analyses assuming GClb is administered over six cycles (list prices; efficacy remains as per CLL14)

Incremental results of VenG vs comparator	Incremental discounted costs	Incremental discounted QALYs	ICER, £ per QALY	Net monetary benefit
Scenario 1: without del(17p)/*TP53* mutation 6 cycles of Clb at a dose used in the CLL14 trial (0.5 mg/kg on Days 1 and 15) and efficacy based on 12 cycles as per CLL14 trial				

Base case: without del(17p)/*TP53* mutation	[redacted]	1.057	[redacted]	[redacted]
Scenario 1: vs GClb	[redacted]	1.057	[redacted]	[redacted]
Scenario 2: without del(17p)/*TP53* mutation 6 cycles of Clb at a dose used in the UK clinical practice dose (10 mg/m^2 on Days 1 and 15) and efficacy based on 12 cycles as per CLL14 trial				
Base case: without del(17p)/*TP53* mutation	[redacted]	1.057	[redacted]	[redacted]
Scenario 2: vs GClb	[redacted]	1.057	[redacted]	[redacted]
Scenario 3: without del(17p)/*TP53* mutation 6 cycles of Clb at a dose used in the UK clinical practice dose (10 mg/m^2 on Days 1 to 7) and efficacy based on 12 cycles as per CLL14 trial				
Base case: without del(17p)/*TP53* mutation	[redacted]	1.057	[redacted]	[redacted]
Scenario 3: vs GClb	[redacted]	1.057	[redacted]	[redacted]

Abbreviations: ICER: incremental cost-effectiveness ratio; QALY: quality-adjusted life year; VenG: venetoclax with obinutuzumab.

5.2.11 Model validation and face validity check

The company took a number of reasonable steps to validate the submitted economic model. To ascertain that the model is clinically valid, AbbVie held an advisory board meeting to discuss the model structure, key model assumptions, and associated inputs with clinicians knowledgeable in CLL and health economists. Quality checks were also carried out using a pre-specified model quality check template and two health economist modellers reviewed the model and its underlying assumptions. Challenges in relation to OS extrapolation beyond the CLL14 trial period were presented to a leading CLL clinician involved with the CLL trials. Two clinical experts who had previously participated in the advisory board provided their opinion on the degree to which the outcomes are valid to help determine the external validity of the model extrapolations. The ERG believes that the above activities and approaches to model validation are appropriate.

The ERG assessed the face validity of the model, particularly with respect to suitability of the constructed structure, appropriateness of data sources and inputs, and plausibility of the obtained results. The structure of the submitted model was scrutinised in order to ascertain than no meaningful health states and pathways have been omitted. While the model is parsimonious, the ERG is satisfied that a partitioned survival model is suitable for the particular decision problem and available data, and is in line with the approach taken in previous CLL appraisals. The ERG also notes that important elements of the analysis (e.g. the adopted perspective, time horizon and discount rates) are in agreement with the NICE Reference Case.

The ERG felt that the company took reasonable steps to ascertain that evidence used in the model was rigorous and suitable. Much of the data used to populate key model parameters were obtained from relevant randomised clinical trials, particularly CLL14, and previous technology appraisals. In instances where the choice of evidence was not drawn from the CLL14 trial, the ERG felt that the evidence employed was *per se* largely appropriate (with the exception of 'pre-progression, off treatment' status). In cases where inappropriate use of evidence or errors in the calculations of input values were identified by the ERG (e.g., incorrect estimates of uncertainty around OS and PFS HR for ibrutinib) these have been queried with the company and highlighted in the critique above. The validity of various assumptions incorporated in the analysis (e.g. related to the treatment and progression pathways, information about NHS services and care routinely provided to CLL patients) was scrutinised by seeking expert opinion from the ERG's clinical expert.

The economic model, which was submitted in a spreadsheet, was also scrutinised by the ERG. Wherever possible, 'extreme value' tests were performed, by replacing the base-case value of influential variables with low and high estimates. Results were found to agree with expectations about the direction and magnitude of change in model parameters and final results. Examination of macros (VBA coding) used to perform simulations did not identify errors in the code.

In summary, the ERG believes the steps undertaken by the company to ensure the validity of the model are appropriate. Putting aside limitations in the analysis due to data immaturity and unavailability, the ERG's examinations deem the model's validity to be, on the whole, sound.

5.3 Exploratory and sensitivity analyses undertaken by the ERG

Based on the critique of the submitted economic model, the ERG suggests an amended base-case analysis. The rationale for these amendments has been given alongside the critique provided in Section 5.2 and is summarised below.

5.3.1 The ERG's preferred base-case analysis

Amendments were implemented to reflect the ERG's preferred base-case analysis in both the non-del(17p)/TP53 and the del(17p)/TP53 populations. These related to the following parameters:

- Utility value for 'pre-progression, off treatment' status. The ERG questions the utility value used by the company for non-progressed patients who are not on treatment. This value is higher than the age-adjusted utility value in the general population and contradicts the rationale used for the choice in utility values elsewhere in the CS, which suggests that CLL patients are unlikely to have better quality of life than their non-CLL counterparts in the general population. Thus, the ERG's preferred approach is to cap the utility value for the 'pre-progression, off treatment' status by using the gender-weighted, age-specific value for members of the UK general public (see Section 5.2.7.1).
- Time-to-event parameters and extrapolations. Changes in PFS, TTNT and OS were implemented in order to obtain extrapolations that the ERG considers more plausible and better aligned with the available data (see Section 5.2.6)

The ERG's preferred base-case values and approaches are summarised below in Table 49 and Table 50 for the populations without and with del(17p)/TP53 population, respectively. Results of the ERG base-case analysis are presented in Sections 5.5.1 and 5.5.2.

Table 49: Summary of values and approached used in the ERG's base case analysis for the non-del(17p)/TP53 population.

Parameter	Value in company's base-case analysis	Value in ERG's preferred base-case analysis	Section where justification for amendment is given
Utility value: 'pre-progression, off treatment' status.	0.82	0.77	Section 5.2.7.1
Time-to-event parameters			Section 5.2.6
PFS	Independent log-logistic extrapolation of CLL14 data	Independent 2-knot hazard spline extrapolation of CLL14 data	
TTNT	Independent log-logistic extrapolation of CLL14 data	Hazard ratio between TTNT and PFS calculated from recreated CLL14 data applied to ERG PFS extrapolation.	
OS	Used OS exponential extrapolation of GClb data from CLL14 for both arms	Used exponential model fitted to IPD from ERIC study to extrapolate beyond 3 years from CLL14 data.	

Table 50: Summary of values and approach used in the ERG's base case analysis for the del(17p)/TP53 population

Parameter	Value in company's base case analysis	Value in ERG's preferred base case analysis	Section where justification for amendment is given
Utility value: 'pre-progression, off treatment' status.	0.82	0.77	Section 5.2.7.1
Time-to-event parameters			Section 5.2.6
PFS	Independent log-logistic extrapolation of CLL14 data and hazard ratio for ibrutinib	Independent 1-knot hazard spline extrapolation of CLL14 data and hazard ratio for ibrutinib	
TTNT	Independent log-logistic extrapolation of CLL14 data	Dependent 1 knot hazard spline extrapolation of CLL14 data	

OS	Used OS exponential extrapolation of VenG data from CLL14 and hazard ratio for ibrutinib	Dependent 1-knot hazard spline extrapolation of CLL14 data and hazard ratio for ibrutinib	
TOT	Data from CLL14 for VenG and equal to PFS for ibrutinib	Same as company	

5.3.2 Additional sensitivity analyses undertaken by the ERG

As discussed in Section 5.2.10, the company carried out an extensive range of sensitivity analyses (reported in the original CS and in subsequent answers to ERG's requests for clarifications) which limited the need for additional sensitivity analyses. However, two additional scenario analyses were carried out by the ERG, where the ERG's preferred base-case amendments were altered. These involved:

- Using an alternative utility value for the progression-free, off treatment sub-state. This value was calculated on the basis of the male/female split specific to CLL patients taken from Cancer Research UK incidence statistics[3] and is applicable to both the non-del(17p)/TP53 and del(17p)/TP53 mutation populations. The value used in this sensitivity analysis is 0.773457027 (compared to 0.7703 in the ERG's base-case and 0.820 in the company's base case).
- Carrying out the ERG's preferred amendments in TTE parameters other than the change in OS for VenG, whilst maintaining the utility value of 0.820 for the progression-free off treatment substate. This was intended to explore whether VenG would remain less costly and more effective despite a decrease in the cost of subsequent treatments following first line treatment with GClb. This is applicable to the non-del(17p)/TP53 mutation population.

The results of these analyses are given in Section 5.5.2 below.

5.4 Conclusions of the cost-effectiveness section

Searches in the available literature did not identify any existing economic evaluations that could address the exact decision problem, as detailed in the final scope, for this appraisal. Thus, the company constructed a *de novo* economic model to evaluate the cost-effectiveness of VenG compared to i) GClb (in the non-del(17p)/TP53 mutation population) and ii) ibrutinib (in the del(17p)/TP53 population). A three-state partitioned survival model was presented in the CS and formed the basis for the cost-effectiveness analyses in both populations.

The company's decision problem addressed in the cost-effectiveness is largely consistent with the NICE scope, although there are some deviations related to the population in the CLL14 and exclusion of treatments (see Section 3 for critique and justifications). The analytic elements of the model (including the chosen model structure, time horizon, discounting, evaluation of costs and outcomes) are generally in line with the NICE Guide to Methods of Technology Appraisal and past NICE Technology Appraisals in CLL.[32, 34]

The immaturity of the CLL14 trial data and reliance on an unadjusted naïve indirect comparison (for VenG vs ibrutinib) add a notable layer of uncertainty to time-to-event extrapolations. Time-to-event data are drivers of incremental costs and outcomes in the decision model. Limitations in currently available data make it difficult to draw a complete and reliable picture of each treatment's effectiveness and they inevitably affect the final cost-effectiveness results. The ERG has identified extrapolations that, we believe, are more plausible and appropriate; these have been incorporated in the ERG's preferred base-case analysis.

Employed health state utility values were sourced from the literature (TA343[32]), rather than EQ-5D-3L data collected in the CLL14 trial. The justification for not using CLL14 trial—that is, the unexpectedly high EQ-5D values that exceed those of the general age-adjusted population—is considered to be reasonable. QALY decrements due to adverse events were appropriately applied. However, the ERG considers the utility value assigned to reflect the 'progression-free, off treatment' status to be problematic. An alternative value has been put forward as a more plausible estimate in the ERG's base-case analysis.

A number of NHS services and their relevant costs were identified and taken into account in cost calculations. These included acquisition and administration costs for first and second line treatments, routine care and tests, cost of TLS prophylaxis and terminal care costs. Cost components included and analytic methods used in the cost calculations are, generally, in line with previous technology appraisals in CLL.

In the company's preferred base-case analysis, and on the basis of list prices for all treatments, VenG is associated with a greater number of QALYs and lower costs against its comparator in the non-del(17p)/TP53 mutation population, suggesting that VenG is dominant versus GClb. In the del(17p)/TP53 mutation population, VenG resulted in a lower average number of QALYs and lower costs versus ibrutinib. In both cases, VenG resulted in a large, positive NMB at both the £20,000 and £30,000 per additional QALY willingness to pay thresholds. The ERG has

undertaken additional comparisons using confidential discounted prices for all treatments; results are reported in an accompanying confidential addendum.

A range of sensitivity analyses, in the form of probabilistic sensitivity analysis, deterministic sensitivity analyses (where parameter values were replaced by low and high estimates) and scenario analyses (where alternative plausible values and approaches were tested) were carried out to explore the robustness of these findings. Additional scenario analyses were also presented in response to requests from the ERG (including for a six-cycle treatment schedule for Clb, utility values derived from CLL14 and use of the CSR mutation status classification in the economic model). Overall, results are relatively robust to changes, supporting the suggestion that, given the existing state of evidence and at typically acceptable WTP threshold values, VenG is a cost-effective option in both the non-del(17p)/TP53 and del(17p)/TP53 mutation populations.

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

5.5.1 Results of the ERG's preferred base-case

The effect of the ERG's preferred base-case amendments on the cost-effectiveness results for the non-del(17p)/TP53 and del(17p)/TP53 populations are reported below. For ease of interpretation, the emphasis on cost-effectiveness estimates are placed on NMB (at a £30,000 WTP threshold and, additionally at a £20,000 WTP threshold), rather than ICERs, as results fall within quadrants of the cost-effectiveness plane where ICER values are not informative or require a different interpretation. Positive values of NMB indicating that a treatment is cost-effective at a given willingness-to-pay threshold; zero indicates equivalence and negative values indicate that a treatment is not cost-effective at the particular threshold. Presented results are calculated based on list prices and percentage changes are calculated on the basis of NMB at a £30,000 WTP per additional QALY.

5.5.1.1 ERG's base-case results for the non-del(17p)/TP53 mutation population.

The effect of the ERG's base-case amendments on the results for the non-del(17p)/TP53 mutation population, when each change is carried out one at a time, can be seen in Table 51.

A change to ERG's preferred utility value for the 'pre-progression, off treatment' status resulted in a small reduction in incremental QALYs (VenG vs GClb), leading to a decrease in the NMB by about [redacted] compared to the company's base-case results. After implementing this adjustment, the revised NMB was found to be [redacted] and [redacted] at WTP thresholds of £20,000 and £30,000 per QALY, respectively.

The effect of the ERG's changes in TTE parameters and extrapolations was more prominent. Carrying out these changes resulted in a sizeable reduction in both the difference in costs and the difference in QALYs between VenG and GClb, leading to a decrease by nearly [redacted] in the company's base case findings. In this case, the revised NMB was found to be [redacted] and [redacted] at WTP thresholds of £20,000 and £30,000 per QALY, respectively.

Table 51. Results for the non-del(17p)/TP53 mutation population when ERG's amendments are implemented one at a time (at list prices).

Treatment	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	ICER (£/QALY)	NMB (WTP: £20k per QALY)	NMB (WTP: £30k per QALY)
Company's base-case results							
GClb	[redacted]	6.742				[redacted]	[redacted]
VenG	[redacted]	7.799	[redacted]	1.057	[redacted]		
Results based on ERG's change in utility value for 'pre-progression, off treatment' status only							
GClb	[redacted]	6.601				[redacted]	[redacted]
VenG	[redacted]	7.418	[redacted]	0.818	[redacted]		
Results based on ERG's change in time-to-event parameters and extrapolations only							
GClb	[redacted]	5.692				[redacted]	[redacted]
VenG	[redacted]	6.279	[redacted]	0.587	[redacted]		

Abbreviations: GClb: chlorambucil with obinutuzumab; ICER: incremental cost effectiveness ratio; QALYs: quality-adjusted life years; VenG: venetoclax with obinutuzumab; NMB: net monetary benefit; WTP: willingness to pay.

As anticipated, carrying out these ERG amendments simultaneously—that is, implementing the ERG's suggested base-case analysis—resulted in reductions in incremental costs and QALYs compared to the company's base-case values, leading to an overall reduction in NMB by approximately [redacted]. The resulting ERG's base-case NMB in the non-del(17p)/TP53 mutation population were [redacted] and [redacted] at WTP thresholds of £20,000 and £30,000 per QALY, respectively (Table 52).

Table 52. ERG's base case results for the non-del(17p)/TP53 mutation population (at list prices)

Treatment	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	ICER (£/QALY)	NMB (WTP: £20k per QALY)	NMB (WTP: £30k per QALY)
Company's preferred base-case							
GClb	[redacted]	6.742				[redacted]	[redacted]
VenG	[redacted]	7.799	[redacted]	1.057	[redacted]		
ERG's preferred base-case							
GClb	[redacted]	5.572				[redacted]	[redacted]
VenG	[redacted]	6.027	[redacted]	0.454	[redacted]		

Abbreviations: GClb: chlorambucil with obinutuzumab; ICER: incremental cost effectiveness ratio; QALYs: quality-adjusted life years; VenG: venetoclax with obinutuzumab; NMB: net monetary benefit; WTP: willingness to pay.

5.5.1.2 ERG's base-case results for the del(17p)/TP53 mutation population.

The effect of the ERG's base-case amendments on the results for the del(17p)/TP53 mutation population, when each change is carried out one at a time, can be seen in Table 53 below.

Using the ERG's preferred utility value for 'pre-progression, off treatment' resulted in a reduction in incremental QALYs (VenG vs ibrutinib), leading to a decrease in the NMB by about [redacted] compared to the company's base-case results. After implementing this adjustment, the revised NMB was found to be [redacted] and [redacted] at WTP thresholds of £20,000 and £30,000 per QALY, respectively.

The ERG's changes in TTE specifications resulted in a sizeable reduction in the difference in costs and a modest reduction on the difference in QALYs between VenG and GClb, leading to a decrease in NMB by nearly [redacted] compared to the company's base case findings. In this case, the revised NMB was found to be [redacted] and [redacted] at WTP thresholds of £20,000 and £30,000 per QALY, respectively.

Table 53: Results for the del(17p)/TP53 mutation population when ERG's amendments are implemented one at a time (at list prices).

Treatment	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	ICER (£/QALY)	NMB (WTP: £20k per QALY)	NMB (WTP: £30k per QALY)
Company's preferred base-case							
Ibrutinib	[redacted]	4.153				[redacted]	[redacted]
VenG	[redacted]	3.991	[redacted]	-0.163	[redacted]		
Results based on ERG's change in utility value for 'pre-progression, off treatment' status only							

Ibrutinib	[redacted]	4.153				[redacted]	[redacted]
VenG	[redacted]	3.802	[redacted]	-0.351	[redacted]		
Results based on ERG's change in time-to-event parameters and extrapolations only							
Ibrutinib	[redacted]	3.690				[redacted]	[redacted]
VenG	[redacted]	3.451	[redacted]	-0.238	[redacted]		
Abbreviations: ICER: incremental cost effectiveness ratio; QALYs: quality-adjusted life years; VenG: venetoclax with obinutuzumab; NMB: net monetary benefit; WTP: willingness to pay. *This ICER falls within the south west quadrant of a cost-effectiveness plane, denoting cost savings per QALY forgone.							

Implementing the ERG's suggested base-case analysis—that is, carrying out all ERG amendments simultaneously—resulted in reductions in incremental costs (cost savings) and QALYs compared to the company's base-case values, leading to an overall reduction in NMB by approximately [redacted]. The resulting ERG's base-case NMB in the del(17p)/TP53 mutation population were [redacted] and [redacted] at WTP thresholds of £20,000 and £30,000 per QALY, respectively (Table 54). The ICER for this comparison falls within the south west quadrant of the cost-effectiveness plane reflecting cost savings per QALY forgone. The ICER resulting from the ERG base-case was [redacted] per QALY, as opposed to the company's base-case ICER at [redacted] per additional QALY.

Table 54: ERG's base case results for the del(17p)/TP53 mutation population (at list prices)

Treatment	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	ICER (£/QALY)	NMB (WTP: £20k per QALY)	NMB (WTP: £30k per QALY)
Company's preferred base-case							
Ibrutinib	[redacted]	4.153				[redacted]	[redacted]
VenG	[redacted]	3.991	[redacted]	-0.163	[redacted]		
ERG's preferred base-case							
Ibrutinib	[redacted]	3.690				[redacted]	[redacted]
VenG	[redacted]	3.326	[redacted]	-0.363	[redacted]		
Abbreviations: ICER: incremental cost effectiveness ratio; QALYs: quality-adjusted life years; VenG: venetoclax with obinutuzumab; NMB: net monetary benefit; WTP: willingness to pay. *This ICER falls within the south west quadrant of a cost-effectiveness plane, denoting cost savings per QALY forgone.							

5.5.2 Additional sensitivity analyses carried out by the ERG.

Results of the additional sensitivity analyses run by the ERG can be seen below. A description of the analysis is given in Section 5.3.2. Briefly, the analysis involved: (i) using an alternative utility value for 'progression-free, off treatment' status (Scenario 1 in Table 55) and (ii) carrying out

the ERG's preferred amendments in TTE parameters but keeping the OS for VenG as per the company's specifications (Scenario 2 in Table 55). Under either scenario, findings agreed in direction with the results of the company's and the ERG's base-case analyses.

Table 55: Additional analyses carried out by the ERG.

Treatment	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	ICER (£/QALY)	NMB (WTP: £20k per QALY)	NMB (WTP: £30k per QALY)
Scenario 1: alternative utility value for 'progression-free, off treatment' status (non-del(17p)/TP53 mutation population)							
GClb	[redacted]	6.610				[redacted]	[redacted]
VenG	[redacted]	7.443	[redacted]	0.833	[redacted]		
Scenario 1: alternative utility value for 'progression-free, off treatment' status (del(17p)/TP53 mutation population)							
Ibrutinib	[redacted]	4.153				[redacted]	[redacted]
VenG	[redacted]	3.814	[redacted]	-0.339	[redacted]		
Scenario 2: No change in OS for VenG (non-del(17p)/TP53 mutation population)							
GClb	[redacted]	5.692				[redacted]	[redacted]
VenG	[redacted]	7.232	[redacted]	1.541	[redacted]		

Abbreviations: GClb: chlorambucil with obinutuzumab; ICER: incremental cost effectiveness ratio; QALYs: quality-adjusted life years; VenG: venetoclax with obinutuzumab; NMB: net monetary benefit WTP: willingness to pay.
*This ICER falls within the south west quadrant of a cost-effectiveness plane, denoting cost savings per QALY forgone.

6 End of life

The CS does not comment on the NICE end of life criteria in relation to VenG. The ERG considers this appropriate as the untreated CLL population would not normally have a life expectancy of less than 24 months when starting treatment with VenG.

7 Innovation

The CS considers that the innovative potential of VenG is demonstrated with the evidence from the CLL14 trial. The company cites the efficacy across all trial subgroups and the manageable adverse event profile along with high rates of undetectable MRD. In addition, the CS says that VenG provides a greater range of treatment options for the unfit CLL population, that VenG avoids the need for chemo-immunotherapy and that because of the fixed treatment duration VenG enables many patients to experience time without therapy. The CS states that this reduces the overall cost burden of treatment in this patient group. The ERG's clinical expert agrees that VenG is innovative, as targeted therapy avoiding traditional chemotherapy has not previously been considered for first-line treatment.

8 Overall conclusions

8.1 Clinical effectiveness evidence

Although there is good quality evidence for the effectiveness of VenG compared with GClb, the ERG has concerns regarding the maturity of the data and the generalisability to the UK population. The comparison of VenG to ibrutinib in the subgroup of people with del(17p)/TP53 mutation is associated with a high level of uncertainty meaning no conclusion of superiority can be made.

8.2 Cost-effectiveness evidence

The economic analysis carried out by the company is, on the whole, appropriate. Given the existing state of evidence, VenG appears to be a cost-effective option at conventional WTP thresholds in both the non-del(17p)/TP53 and del(17p)/TP53 mutation populations. However, immaturity of key effectiveness data (in VenG vs. GClb) and reliance on an unadjusted naïve

indirect comparison (for VenG vs ibrutinib) inevitably affect the cost-effectiveness calculations and introduce a layer of uncertainty in the overall results.

9 References

1. Te Raa GD, Kater AP. TP53 dysfunction in CLL: Implications for prognosis and treatment. *Best Pract Res Clin Haematol* 2016;**29**(1):90-9. http://dx.doi.org/10.1016/j.beha.2016.08.002
2. Schuh AH, Parry-Jones N, Appleby N, Bloor A, Dearden CE, Fegan C*, et al.* Guideline for the treatment of chronic lymphocytic leukaemia: A British Society for Haematology Guideline. *Br J Haematol* 2018;**182**(3):344-59. http://dx.doi.org/10.1111/bjh.15460
3. Cancer Research UK. *Chronic lymphocytic leukaemia (CLL) incidence statistics.* . London: Cancer Research UK. URL: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/leukaemia-cll/incidence (Accessed 13 June 2019).
4. Shanafelt TD, Bowen D, Venkat C, Slager SL, Zent CS, Kay NE*, et al.* Quality of life in chronic lymphocytic leukemia: an international survey of 1482 patients. *Br J Haematol* 2007;**139**(2):255-64. http://dx.doi.org/10.1111/j.1365-2141.2007.06791.x
5. Leukaemia Care. *Chronic Lymphocytic Leukaemia (CLL): A Guide for Patients*. Worcester: Leukaemia Care; 2019. URL: https://www.leukaemiacare.org.uk/wp-content/uploads/Chronic-Lymphocytic-Leukaemia-CLL-Web-Version.pdf (Accessed 13 June, 2019).
6. Hanada M, Delia D, Aiello A, Stadtmauer E, Reed JC. bcl-2 gene hypomethylation and high-level expression in B-cell chronic lymphocytic leukemia. *Blood* 1993;**82**(6):1820-8.
7. Fischer K, Al-Sawaf O, Bahlo J, Fink A-M, Tandon M, Dixon M*, et al.* Venetoclax and Obinutuzumab in Patients with CLL and Coexisting Conditions. *New England Journal of Medicine* 2019;**380**(23):2225-36. http://dx.doi.org/10.1056/NEJMoa1815281
8. Skoetz N, Bauer K, Elter T, Monsef I, Roloff V, Hallek M*, et al.* Alemtuzumab for patients with chronic lymphocytic leukaemia. *Cochrane Database of Systematic Reviews* 2012(2).
9. Flinn IW, Gribben JG, Dyer MJS, Wierda W, Maris MB, Furman RR*, et al.* Phase 1b study of venetoclax-obinutuzumab in previously untreated and relapsed/refractory chronic lymphocytic leukemia. *Blood* 2019;**133**(26):2765-75. http://dx.doi.org/10.1182/blood-2019-01-896290
10. AbbVie Inc. Study BO25323 (CLL14) Primary Clinical Study Report. Data on File. 2019.
11. Hallek M, Cheson BD, Catovsky D, Caligaris-Cappio F, Dighiero G, Dohner H*, et al.* iwCLL guidelines for diagnosis, indications for treatment, response assessment, and supportive management of CLL. *Blood* 2018;**131**(25):2745-60. http://dx.doi.org/10.1182/blood-2017-09-806398
12. Fischer K, Al-Sawaf O, Fink A-M, Dixon M, Bahlo J, Warburton S*, et al.* Venetoclax and obinutuzumab in chronic lymphocytic leukemia. *Blood* 2017;**129**(19):2702-5. http://dx.doi.org/10.1182/blood-2017-01-761973
13. ClinicalTrials.gov. *A Phase 1 Study Evaluating the Safety and Pharmacokinetics of ABT-199 in Subjects With Relapsed or Refractory Chronic Lymphocytic Leukemia and Non-Hodgkin Lymphoma.* Bethseda: National Library of Medicine; 2019 (last updated). URL: https://clinicaltrials.gov/ct2/show/NCT01328626?term=M12-175&rank=1 (Accessed 28 June, 2019).
14. Dohner H, Stilgenbauer S, Benner A, Leupolt E, Krober A, Bullinger L*, et al.* Genomic aberrations and survival in chronic lymphocytic leukemia. *N Engl J Med* 2000;**343**(26):1910-6. http://dx.doi.org/10.1056/nejm200012283432602
15. Julia V, Carsten N, Kater AP, Bahlo J, Farstenau M, Fink AM*, et al.* The GAIA (CLL13) trial: An international intergroup phase III study for frontline therapy in chronic lymphocytic leukemia (CLL). *Journal of Clinical Oncology* 2018;**36**(15). http://dx.doi.org/10.1200/JCO.2018.36.15_suppl.TPS7582
16. Owen C, Christofides A, Johnson N, Lawrence T, MacDonald D, Ward C. Use of minimal residual disease assessment in the treatment of chronic lymphocytic leukemia. *Leuk Lymphoma* 2017;**58**(12):2777-85. http://dx.doi.org/10.1080/10428194.2017.1318439
17. European Medicines Agency. *Appendix 4 to the Guideline on the valuation of anticancer medicinal products in man. EMA/CHMP/703715/2012 Rev. 2* Amsterdam: European Medicines

Agency; 2015. URL: https://www.ema.europa.eu/en/appendix-4-guideline-evaluation-anticancer-medicinal-products-man-condition-specific-guidance (Accessed 18 July, 2019).
18. Thompson M, Brander D, Nabhan C, Mato A. Minimal Residual Disease in Chronic Lymphocytic Leukemia in the Era of Novel Agents: A Review. *JAMA Oncology* 2018;**4**(3):394-400. http://dx.doi.org/10.1001/jamaoncol.2017.2009
19. Food and Drug Administration. *NDA/BLA Multi-disciplinary Review and Evaluation Supplemental NDA 208573 S-13 VENCLEXTA (venetoclax)*. Maryland: Food and Drug Administration; 2019; https://www.accessdata.fda.gov/drugsatfda_docs/nda/2019/208573Orig1s013MultidisciplineR.pdf). http://dx.doi.org/https://www.accessdata.fda.gov/drugsatfda_docs/nda/2019/208573Orig1s013MultidisciplineR.pdf
20. Mato AR, Roeker LE, Allan JN, Pagel JM, Brander DM, Hill BT*, et al.* Outcomes of front-line ibrutinib treated CLL patients excluded from landmark clinical trial. *Am J Hematol* 2018;**93**(11):1394-401. http://dx.doi.org/10.1002/ajh.25261
21. Ahn IE, Farooqui MZH, Tian X, Valdez J, Sun C, Soto S*, et al.* Depth and durability of response to ibrutinib in CLL: 5-year follow-up of a phase 2 study. *Blood* 2018;**131**(21):2357-66. http://dx.doi.org/10.1182/blood-2017-12-820910
22. Woyach JA, Ruppert AS, Heerema NA, Zhao W, Booth AM, Ding W*, et al.* Ibrutinib Regimens versus Chemoimmunotherapy in Older Patients with Untreated CLL. *N Engl J Med* 2018;**379**(26):2517-28. http://dx.doi.org/10.1056/NEJMoa1812836
23. Burger JA, Tedeschi A, Barr PM, Robak T, Owen C, Ghia P*, et al.* Ibrutinib as Initial Therapy for Patients with Chronic Lymphocytic Leukemia. *New England Journal of Medicine* 2015;**373**(25):2425-37. http://dx.doi.org/10.1056/NEJMoa1509388
24. Farooqui MZ, Valdez J, Martyr S, Aue G, Saba N, Niemann CU*, et al.* Ibrutinib for previously untreated and relapsed or refractory chronic lymphocytic leukaemia with TP53 aberrations: a phase 2, single-arm trial. *Lancet Oncol* 2015;**16**(2):169-76. http://dx.doi.org/10.1016/s1470-2045(14)71182-9
25. National Institute for Health and Care Excellence (NICE). *Ibrutinib for previously treated chronic lymphocytic leukaemia and untreated chronic lymphocytic leukaemia with 17p deletion or TP53 mutation. Technology appraisal guidance [TA429]*. National Institute for Health and Care Excellence: London; 2017. URL: https://www.nice.org.uk/guidance/ta429 (Accessed 13 June, 2019).
26. Herishanu Y, Shaulov A, Fineman R, Bašić‐Kinda S, Aviv A, Wasik‐Szczepanek E*, et al.* Frontline treatment with the combination obinutuzumab ± chlorambucil for chronic lymphocytic leukemia outside clinical trials: Results of a multinational, multicenter study by ERIC and the Israeli CLL study group. *American Journal of Hematology* 2020; 10.1002/ajh.25766. http://dx.doi.org/10.1002/ajh.25766
27. Al-Sawaf O, Gentile B, Devine J, Sail K, Tandon M, Fink A-M*, et al.* Rapid Improvement of Patient-Reported Outcomes with Venetoclax Plus Obinutuzumab in Patients with Previously Untreated CLL and Coexisting Conditions: A Prospective Analysis from the CLL14 Trial. Paper presented at: 61st ASH Annual Meeting: Oral and Poster Abstracts; Orlando.
28. Fischer K, Al-Sawaf O, Bahlo J, Fink AM, Tandon M, Dixon M*, et al.* Effect of fixed-duration venetoclax plus obinutuzumab (VenG) on progression-free survival (PFS), and rates and duration of minimal residual disease negativity (MRD-) in previously untreated patients (pts) with chronic lymphocytic leukemia (CLL) and comorbidities. *Journal of Clinical Oncology* 2019;**37**(15).
29. Fischer K, Porro Lura M, Al-Sawaf O, Bahlo J, Fink A, Tandon M*, et al.* Fixed-duration venetoclax plus obinutuzumab improves pfs and minimal residual disease negativity in patients with previously untreated CLL and comorbidities. *Hematological Oncology* 2019;**37 (Supplement 2)**:82-4. http://dx.doi.org/http://dx.doi.org/10.1002/hon.52_2629
30. Fischer K, Ritgen M, Al-Sawaf O, Robrecht S, Tandon M, Fink A-M*, et al.* Quantitative Analysis of Minimal Residual Disease (MRD) Shows High Rates of Undetectable MRD after Fixed-Duration Chemotherapy-Free Treatment and Serves As Surrogate Marker for Progression-Free Survival: A

Prospective Analysis of the Randomized CLL14 Trial. *Blood* 2019;**134**(Supplement_1):36-. http://dx.doi.org/10.1182/blood-2019-125825
31. Goede V, Fischer K, Busch R, Engelke A, Eichhorst B, Wendtner CM*, et al.* Obinutuzumab plus Chlorambucil in Patients with CLL and Coexisting Conditions. *New England Journal of Medicine* 2014;**370**(12):1101-10. http://dx.doi.org/10.1056/NEJMoa1313984
32. National Institute for Health and Care Excellence (NICE). *Obinutuzumab in combination with chlorambucil for untreated chronic lymphocytic leukaemia. Technology appraisal guidance [TA343].* . London: National Institute for Health and Care Excellence; 2015. URL: https://www.nice.org.uk/Guidance/TA343 (Accessed 12 June, 2019).
33. Office for National Statistics. National life tables, UK: 2015 to 2017. https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/lifeexpectancies/bulletins/nationallifetablesunitedkingdom/2015to2017. Accessed 03-Oct-2019.
34. National Institute for Health and Care Excellence (NICE). *Venetoclax with rituximab for previously treated chronic lymphocytic leukaemia. Technology appraisal guidance [TA561].* . London: National Institute for Health and Care Excellence; 2019. URL: https://www.nice.org.uk/guidance/ta561 (Accessed 03 June, 2019).
35. Shanafelt TD, Byrd JC, Call TG, Zent CS, Kay NE. Narrative Review: Initial Management of Newly Diagnosed, Early-Stage Chronic Lymphocytic LeukemiaInitial Management of Newly Diagnosed, Early-Stage Chronic Lymphocytic Leukemia. *Annals of Internal Medicine* 2006;**145**(6):435-47. http://dx.doi.org/10.7326/0003-4819-145-6-200609190-00007
36. Wei Y, Royston P. Reconstructing time-to-event data from published Kaplan-Meier curves. *Stata J* 2017;**17**(4):786-802.
37. Hillmen P, Janssens A, Babu KG, Kloczko J, Grosicki S, Manson S*, et al.* Health-related quality of life and patient-reported outcomes of ofatumumab plus chlorambucil versus chlorambucil monotherapy in the COMPLEMENT 1 trial of patients with previously untreated CLL. *Acta Oncologica* 2016;**55**(9-10):1115-20.
38. Mato AR, Allan JN, Pagel JM, Brander DM, Hill BT, Cheson BD*, et al.* Front-Line Ibrutinib Therapy for Chronic Lymphocytic Leukemia (CLL) in the Real World: Responses, Toxicity, Outcomes and Subsequent Therapies. *Blood* 2017;**130**(Suppl 1):3011.
39. Dolan P. Modeling Valuations for EuroQol Health States. *Medical Care* 1997;**35**(11):1095-108. http://dx.doi.org/10.1097/00005650-199711000-00002
40. Kay NE, Flowers CR, Weiss M, Lamanna N, Flinn IW, Grinblatt D*, et al.* Variation in health-related quality of life by line of therapy of patients with chronic lymphocytic leukemia. In: Am Soc Hematology; 2012.
41. Pashos CLF, Christopher R. Weiss, Mark Lamanna, Nicole Farber, Charles M. Kipps, Thomas J. Lerner, Susan Kay, Neil Sharman, Jeff Porte Grinblatt, David L. Flinn, Ian W. Kozloff, Mark Swern, Arlene S. Kahn, Zeba M. Street, Thomas K. Sullivan, Kristen A. Harding, Gale Keating, Michael J. Variation in health-related quality of life by age among patients with chronic lymphocytic leukemia. *Blood* 2012;**118**(21).
42. Pashos CL, Flowers CR, Kay NE, Weiss M, Lamanna N, Farber C*, et al.* Association of health-related quality of life with gender in patients with B-cell chronic lymphocytic leukemia. *Supportive Care in Cancer* 2013;**21**(10):2853-60.
43. Ara R, Brazier JE. Populating an economic model with health state utility values: moving toward better practice. *Value Health* 2010;**13**(5):509-18. http://dx.doi.org/10.1111/j.1524-4733.2010.00700.x
44. National Institute for Clinical Excellence (NICE). *Pixantrone monotherapy for treating multiply relapsed or refractory aggressive non-Hodgkin's B-cell lymphoma. Technology appraisal guidance [TA306].* London: National Institute of Health and Care Excellence; 2014. URL: https://www.nice.org.uk/guidance/ta306 (Accessed 23 August, 2019).
45. Lloyd A, Nafees B, Narewska J, Dewilde S, Watkins J. Health state utilities for metastatic breast cancer. *Br J Cancer* 2006;**95**(6):683-90. http://dx.doi.org/10.1038/sj.bjc.6603326

46. National Institute for Health and Care Excellence (NICE). *Bendamustine for the first-line treatment of chronic lymphocytic leukaemia. Technology appraisal guidance [TA216]*. London: National Institute for Health and Care Excellence; 2011. URL: https://www.nice.org.uk/guidance/ta216 (Accessed 17 June, 2019).
47. Beusterien KM, Davies J, Leach M, Meiklejohn D, Grinspan JL, O'Toole A*, et al.* Population preference values for treatment outcomes in chronic lymphocytic leukaemia: a cross-sectional utility study. *Health Qual Life Outcomes* 2010;**8**:50. http://dx.doi.org/10.1186/1477-7525-8-50
48. National Institute for Health and Care Excellence (NICE). *Ofatumumab in combination with chlorambucil or bendamustine for untreated chronic lymphocytic leukaemia. Technology appraisal guidance [TA344].* . London: National Institute for Health and Care Excellence; 2015. URL: https://www.nice.org.uk/guidance/ta344. (Accessed 03 June, 2019).
49. Nafees B, Stafford M, Gavriel S, Bhalla S, Watkins J. Health state utilities for non small cell lung cancer. *Health Qual Life Outcomes* 2008;**6**:84. http://dx.doi.org/10.1186/1477-7525-6-84
50. Tolley K, Goad C, Yi Y, Maroudas P, Haiderali A, Thompson G. Utility elicitation study in the UK general public for late-stage chronic lymphocytic leukaemia. *Eur J Health Econ* 2013;**14**(5):749-59. http://dx.doi.org/10.1007/s10198-012-0419-2
51. National Institute for Health and Care Excellence (NICE). *Idelalisib for treating chronic lymphocytic leukaemia. Technology appraisal guidance [TA359]*. London: National Institute for Health and Care Excellence; 2015. URL: https://www.nice.org.uk/guidance/ta359 (Accessed 17 June, 2019).
52. National Institute for Clinical Excellence (NICE). British National Formulary. https://bnf.nice.org.uk. Accessed 21-Oct-2019.
53. Barr PM, Robak T, Owen C, Tedeschi A, Bairey O, Bartlett NL*, et al.* Sustained efficacy and detailed clinical follow-up of first-line ibrutinib treatment in older patients with chronic lymphocytic leukemia: extended phase 3 results from RESONATE-2. *Haematologica* 2018;**103**(9):1502-10. http://dx.doi.org/10.3324/haematol.2018.192328
54. 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* 2008;**11**(6):A483. http://dx.doi.org/10.1016/S1098-3015(10)66610-7
55. NHS Improvement. *Reference Costs*. 2018. URL: https://improvement.nhs.uk/resources/reference-costs/ (Accessed 21 August, 2019).
56. Cheson BD, Heitner Enschede S, Cerri E, Desai M, Potluri J, Lamanna N*, et al.* Tumor Lysis Syndrome in Chronic Lymphocytic Leukemia with Novel Targeted Agents. *The Oncologist* 2017;**22**(11):1283-91. http://dx.doi.org/10.1634/theoncologist.2017-0055
57. Kater AP, Seymour JF, Hillmen P, Eichhorst B, Langerak AW, Owen C*, et al.* Fixed Duration of Venetoclax-Rituximab in Relapsed/Refractory Chronic Lymphocytic Leukemia Eradicates Minimal Residual Disease and Prolongs Survival: Post-Treatment Follow-Up of the MURANO Phase III Study. *J Clin Oncol* 2019;**37**(4):269-77. http://dx.doi.org/10.1200/jco.18.01580
58. O'Brien S, Furman RR, Coutre S, Flinn IW, Burger JA, Blum K*, et al.* Single-agent ibrutinib in treatment-naive and relapsed/refractory chronic lymphocytic leukemia: a 5-year experience. *Blood* 2018;**131**(17):1910-9. http://dx.doi.org/10.1182/blood-2017-10-810044
59. Davids MS, Hallek M, Wierda W, Roberts AW, Stilgenbauer S, Jones JA*, et al.* Comprehensive Safety Analysis of Venetoclax Monotherapy for Patients with Relapsed/Refractory Chronic Lymphocytic Leukemia. *Clin Cancer Res* 2018;**24**(18):4371-9. http://dx.doi.org/10.1158/1078-0432.ccr-17-3761
60. National Institute for Health and Care Excellence (NICE). *Lenvatinib with everolimus for previously treated advanced renal cell carcinoma [TA498].* . London: National Institute for Health and Care Excellence (NICE); 2018. URL: https://www.nice.org.uk/guidance/ta498. (Accessed 22 October, 2019).

61. Woods B, Hawkins N, Dunlop W, O'Toole A, Bramham-Jones S. Bendamustine versus chlorambucil for the first-line treatment of chronic lymphocytic leukemia in England and Wales: a cost-utility analysis. *Value Health* 2012;**15**(5):759-70. http://dx.doi.org/10.1016/j.jval.2012.03.1389
62. Curtis L. BA. *PSSRU Unit Costs of Health and Social Care 2018. https://kar.kent.ac.uk/70995/. Accessed 15-Apr-2019.*; 2018.
63. Seymour JF, Kipps TJ, Eichhorst B, Hillmen P, D'Rozario J, Assouline S*, et al.* Venetoclax–Rituximab in Relapsed or Refractory Chronic Lymphocytic Leukemia. *New England Journal of Medicine* 2018;**378**(12):1107-20. http://dx.doi.org/10.1056/NEJMoa1713976
64. Round J, Jones L, Morris S. Estimating the cost of caring for people with cancer at the end of life: A modelling study. *Palliat Med* 2015;**29**(10):899-907. http://dx.doi.org/10.1177/0269216315595203