




Lenalidomide with Rituximab for Previously Treated Follicular Lymphoma and Marginal Zone Lymphoma: An Evidence Review Group Perspective of a NICE Single Technology Appraisal

Willem J. A. Witlox¹ · Sabine E. Grimm¹ · Rob Riemsma² · Nigel Armstrong² · Steve Ryder² · Steven Duffy² · Vanesa Huertas Carrera² · Pawel Posadzki² · Gillian Worthy² · Xavier G. L. V. Pouwels^{1,4} · Bram L. T. Ramaekers¹ · Jos Kleijnen^{2,3} · Manuela A. Joore^{1,3} · Antoinette D. I. van Asselt^{1,5,6} 

Accepted: 15 October 2020
© The Author(s) 2020

Abstract

The National Institute for Health and Care Excellence (NICE) invited the manufacturer (Celgene) of lenalidomide (Revlimid[®]), as part of the Single Technology Appraisal (STA) process, to submit evidence for the clinical effectiveness and cost-effectiveness of lenalidomide in combination with rituximab (MabThera[®]), together referred to as R^2 , for the treatment of adults with treated follicular lymphoma (FL) or marginal zone lymphoma (MZL). Kleijnen Systematic Reviews Ltd, in collaboration with Maastricht University Medical Centre+, was commissioned to act as the independent Evidence Review Group (ERG). This paper summarises the company submission (CS), presents the ERG's critical review on the clinical and cost-effectiveness evidence in the CS, highlights the key methodological considerations, and describes the development of the NICE guidance by the Appraisal Committee. The CS included one relevant study, for the comparison of R^2 versus rituximab monotherapy (R-mono): the AUGMENT trial. In addition, the company performed an unanchored indirect comparison of R^2 versus rituximab combined with cyclophosphamide, doxorubicin, vincristine, and prednisolone (R-CHOP) and rituximab combined with cyclophosphamide, vincristine, and prednisolone (R-CVP), using data for R^2 from the AUGMENT trial and pooled data for R-CHOP/R-CVP from the Haematological Malignancy Research Network (HMRN) database. During the STA process, the company provided an addendum containing evidence on only the FL population, in line with the marketing authorisation obtained at that time, which did not include MZL. The probabilistic incremental cost-effectiveness ratios (ICERs) presented by the company were £27,768 per quality-adjusted life year (QALY) gained for R^2 versus R-CHOP, £41,602 per QALY gained for R^2 versus R-CVP, and £23,412 per QALY gained for R^2 versus R-mono. The ERG's concerns included the validity of the unanchored comparison, the unavailability of a state transition model to verify the outcomes of the partitioned survival model, substantial uncertainty in survival curves, and potential over-estimation of utility values. The revised ERG base case resulted in ICERs ranging from £16,874 to £44,888 per QALY gained for R^2 versus R-CHOP, from £23,135 to £59,810 per QALY gained for R^2 versus R-CVP, and from £18,779 to £27,156 per QALY gained for R^2 versus R-mono. Substantial uncertainty remained around these ranges. NICE recommended R^2 within its marketing authorisation, as an option for previously treated FL (grade 1–3A) in adults, contingent on the company providing lenalidomide according to the commercial arrangement.

1 Introduction

Lenalidomide, trade name Revlimid[®], in combination with rituximab, trade name MabThera[®] (together abbreviated as R^2), was appraised within the National Institute for Health and Care Excellence (NICE) Single Technology Appraisal (STA) process. Health technologies must be shown to be

clinically effective and to represent a cost-effective use of National Health Service (NHS) resources in order to be recommended by NICE. Within the STA process, the company (Celgene) provided NICE with a written submission and a health economic model, summarising the company's estimates of the clinical effectiveness and cost-effectiveness of R^2 for the treatment of previously treated follicular lymphoma (FL) and marginal zone lymphoma (MZL). This company submission (CS) was reviewed by an Evidence Review Group (ERG) independent of NICE [1]. The ERG, Kleijnen Systematic Reviews in collaboration with

✉ Antoinette D. I. van Asselt
a.d.i.van.asselt@umcg.nl

Extended author information available on the last page of the article

Key Points for Decision Makers

The recommendation in National Institute for Health and Care Excellence (NICE) Technical Support Document 19 to verify extrapolations resulting from a partitioned survival model by means of a state transition model alongside it is rarely to never brought to practice in a Single Technology Appraisal. It is also not necessarily something that committees appreciate having for making decisions. Reconsideration of this recommendation may be called for.

The use of matching-adjusted indirect comparison (MAIC) remains largely untested, and there is a lack of clarity whether the results are relevant to the decision problem. In particular, unanchored MAICs are regarded as unfeasible.

NICE recommended lenalidomide with rituximab, within its marketing authorisation, as an option for previously treated follicular lymphoma (grade 1–3A) in adults, contingent on the company providing lenalidomide according to the commercial arrangement.

Maastricht University Medical Centre+, produced an ERG report [1]. After consideration of the evidence submitted by the company and the ERG report, the NICE Appraisal Consultation Document (ACD) issued guidance whether or not to recommend the technology by means of the Final Appraisal Document (FAD), which is open for appeal. This paper presents a summary of the ERG report and the development of the NICE guidance. Furthermore, it highlights important methodological issues which may help in future decision making.

Full details of all relevant appraisal documents (including the appraisal scope, CS, ERG report, consultee submissions, ACD, FAD, and comments from consultees) can be found on the NICE website [1].

2 The Decision Problem

The CS defined the population as “adults with treated follicular lymphoma or marginal zone lymphoma”, which was in line with the NICE final scope [2]. The intervention (lenalidomide 20 mg orally, with rituximab 375 mg/m² intravenously) and outcomes were also in line with the NICE scope, although the scope did not specify dosages for the intervention and the CS included some additional outcomes (event-free survival, time to next anti-lymphoma and chemotherapy treatment, and response rate to next anti-lymphoma treatment). The comparators in the CS were rituximab in

combination with chemotherapy, and obinutuzumab in combination with bendamustine (O-Benda). This was a deviation from the NICE scope in the sense that O-Benda was not listed as a comparator in the scope, and rituximab monotherapy (R-mono) was, as well as established clinical management without lenalidomide (including, but not limited to, bendamustine).

Following the final marketing authorisation indication for lenalidomide with rituximab [indicated “for the treatment of adult patients with previously treated FL (grade 1–3A)”], the scope of the appraisal focussed on the FL population only.

3 Independent Evidence Review Group (ERG) Review

The ERG reviewed the clinical effectiveness and cost-effectiveness evidence of R^2 for this indication. As part of the STA process, the ERG and NICE had the opportunity to ask for clarification on specific issues in the CS, in response to which the company provided additional information [3]. Based on this information, the ERG produced an ERG base case by modifying the health economic model submitted by the company, and assessed the impact of alternative assumptions and parameter values on the model results. Sections 3.1–3.6 summarise the evidence presented in the CS, as well as the review of the ERG.

3.1 Clinical Effectiveness Evidence Submitted by the company

The CS included six studies that were deemed relevant. Four of these studies evaluated R^2 , of which one was a randomised controlled trial (RCT) of R^2 versus R-mono (the AUGMENT trial) [4]; the remaining three studies did not include relevant comparators according to the NICE scope. A fifth relevant study, by van Oers et al., evaluated rituximab combined with cyclophosphamide, doxorubicin, vincristine, and prednisolone (R-CHOP) versus CHOP [5]. A further study evaluated O-Benda versus bendamustine monotherapy (the GADOLIN trial [6]). These last two studies were included for unanchored indirect comparisons. The trial by van Oers et al. (2006) [5] was used to compare R^2 versus R-CHOP although this study only included rituximab-naïve patients and was therefore not representative for the UK patient population. The GADOLIN study was included for an indirect comparison of R^2 with O-Benda.

The AUGMENT trial was a randomised, double-blind, multicentre, controlled trial comparing R^2 versus R-mono in non-rituximab refractory patients with FL grade 1, 2, or 3A or MZL. The study was conducted across 96 sites in 17 countries outside the UK. Intravenous (IV) rituximab 375 mg/m² was given every week in cycle 1 (days 1, 8, 15,

and 22) and on day 1 of every 28-day cycle for cycles 2–5. R^2 arm patients received lenalidomide once daily on days 1–21 of every 28-day cycle up to 12 cycles. Dose modification rules allowed for dosing down lenalidomide to 2.5 mg. Treatment continued until progression or unacceptable toxicity.

Baseline demographics for the population in the AUGMENT trial were similar between arms. Overall, 261 patients (73%) had Ann Arbor stage III–IV disease; 123 patients (34%) had a Follicular Lymphoma International Prognostic Index (FLIPI) score ≥ 3 ; and 183 patients (51%) had high tumour burden per Group d'Etude des Lymphomes Folliculaires (GELF) criteria.

Results from the AUGMENT trial show favourable results for R^2 when compared to R-mono in terms progression-free survival (PFS) with a greater median PFS (results were confidential). However, there was no evidence of a difference in overall survival (OS) with a hazard ratio of 0.61 (95% confidence interval 0.33–1.13) for patients treated with R^2 compared to R-mono. At the time of the analysis the OS data were immature with 16 deaths on R^2 and 26 deaths on R-mono. Overall response rate (ORR) was significantly greater for R^2 compared with R-mono (78% vs. 53%; $p < 0.0001$). The complete response (CR) rate was also greater for the R^2 arm compared with R-mono (34% vs. 18%; $p = 0.001$). In terms of health-related quality of life, no clinically meaningful change from baseline in the Global Health Status/Quality of Life (GHS/QoL) domain of the European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire-Core30 (QLQ-C30) was observed across any of the post-baseline assessment visits. Between-group differences in mean changes were small and not clinically meaningful across all assessment visits.

Treatment-emergent adverse events (TEAEs) during AUGMENT for the total population (FL and MZL) were reported in 174 patients (99%) in the R^2 arm and 173 patients (96%) in the R-mono arm. More patients in the R^2 arm (69%) experienced a grade 3 or 4 TEAE compared with those in the R-mono arm (32%), and two patients in each treatment arm reported a grade 5 TEAE. Additionally, a greater proportion of patients reported serious adverse events in the R^2 arm (26%) compared with those in the R-mono arm (14%).

The company performed three unanchored indirect comparisons, two using data from published evidence and one using data from the Haematological Malignancy Research Network (HMRN) [7]. The HMRN is a population-based cohort covering the Yorkshire and Humber & Yorkshire Cancer Networks for all patients newly diagnosed with a haematological malignancy between 2004 and 2016.

The unanchored indirect comparisons were as follows:

- R^2 versus R-CHOP for non-rituximab refractory patients, using van Oers et al. (2006) [5] comparing R-CHOP with CHOP (only the R-CHOP arm was used in the analyses).
- R^2 versus O-Benda for rituximab refractory patients, based on comparator data from a study by Sehn et al. (2016) [6] comparing O-Benda with bendamustine monotherapy (only the O-Benda arm was used in the analyses).
- R^2 versus R-CHOP/rituximab combined with cyclophosphamide, vincristine, and prednisolone (R-CVP) for non-rituximab refractory patients. This was done using data from HMRN.

The two unanchored indirect comparisons using published evidence have not been used by the ERG in their deliberations because the study by van Oers et al. is not representative for UK patients, and O-Benda is not a relevant comparator according to the NICE scope.

Results from the remaining matching-adjusted indirect comparison (MAIC) (R^2 versus pooled data for R-CHOP/R-CVP for non-rituximab refractory patients using data from HMRN) show a significant improvement in OS and time to next anti-lymphoma treatment (TTNLT) for R^2 compared with R-CHOP/R-CVP, but no evidence of a difference in PFS. All results were confidential.

3.2 Critique of Clinical Effectiveness Evidence and Interpretation

The CS and response to clarification provided sufficient details for the ERG to appraise the literature searches conducted as part of the systematic review to identify clinical effectiveness studies. A good range of databases and resources were searched.

The CS included one relevant study, for the comparison of R^2 versus R-mono: the AUGMENT trial [4]. All patients in this trial were rituximab refractory. In addition, the company performed an unanchored indirect comparison of R^2 versus R-CHOP and R-CVP, using data for R^2 from the AUGMENT trial and pooled data for R-CHOP/R-CVP from the HMRN database.

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

3.3 Cost-Effectiveness Evidence Submitted by the Company

The company conducted searches for cost-effectiveness, health-related quality of life, and healthcare resource use evidence. Although four economic evaluations from a UK perspective were identified, none included R^2 , and the company therefore chose to base their submission on a de novo cohort-level partitioned survival model (PSM) with three health states: progression-free (PF), post-progression (PP), and death (Fig. 1). The company argued that a PSM was more appropriate than a state transition model (STM) because of a lack of data on post-progression survival (PPS).

The analysis took an NHS and Personal Social Services (PSS) perspective. The model had a time horizon of 40 years with a cycle length of 28 days, and a half-cycle correction was applied. All costs and quality-adjusted life years (QALYs) were discounted at a rate of 3.5% per year.

The patient population considered in the model was in line with the proposed licence: adult patients with previously treated FL or MZL. Due to the similar prognosis of FL and MZL patients, and the difficulty in sourcing MZL-specific data, FL and MZL populations were pooled throughout the economic analysis. After the final marketing authorisation, which did not include MZL, the company provided an addendum containing evidence on only the FL population.

Lenalidomide and rituximab are administered orally and by (IV) infusion, respectively. The comparators in the economic model were rituximab in combination with chemotherapy, i.e. R-CHOP or R-CVP, and O-Benda. The ERG did not include O-Benda in its review, as NICE explicitly stated that it was not considered a relevant comparator for disease that is rituximab refractory.

The main source of evidence on treatment effectiveness used for intervention and comparators was the AUGMENT study [4] for R^2 and HMRN data [7] for R-CHOP and R-CVP.

Based on HMRN data and clinical opinion, the efficacy (OS and PFS) of R-CHOP and R-CVP were assumed to be similar, and hence HMRN data for R-CHOP and R-CVP were pooled. For the economic model, this implied that the comparisons of R^2 versus R-CHOP and R-CVP had identical outcomes for effectiveness (QALYs) and only differed with respect to costs.

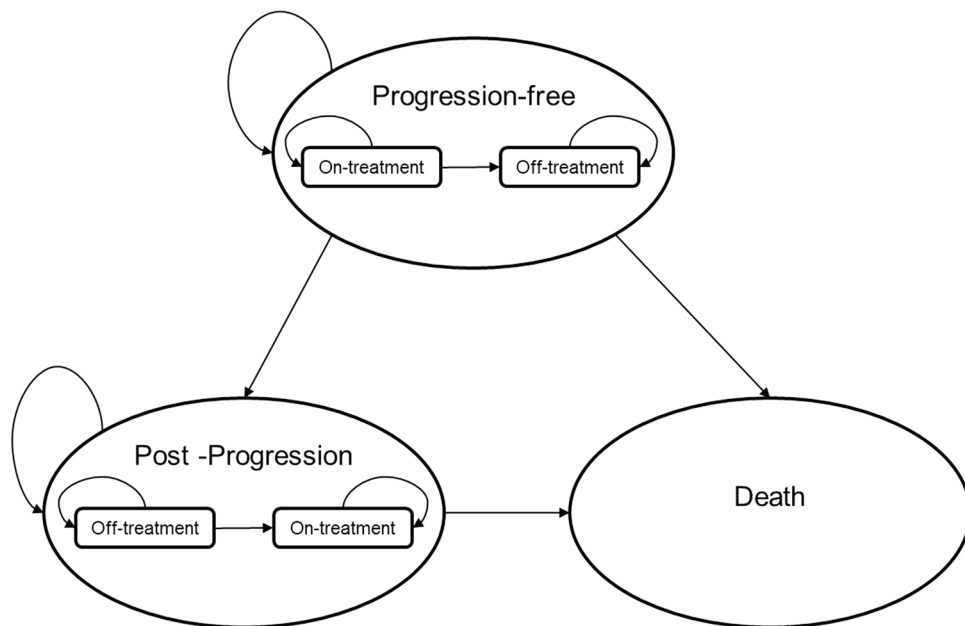
Parametric survival curves were fitted to the matched patient-level data from AUGMENT and HRMN and were then used to extrapolate survival beyond study follow-up. Survival analysis was performed for OS, PFS, TTNLT, and time on treatment (ToT). PFS and ToT data were used to determine the number of patients staying in the PF (on- and off-treatment) health state. The proportion of patients moving to the PP (on- and off-treatment) health state was based on PFS, TTNLT, and OS data. The curves were adjusted for

treatment waning, which in the company's base case was assumed to occur at 5 years, consistent with previous NICE submissions in the same disease area (TA472 [8] and TA137 [9]). After this time point, the comparator hazard of progressing or dying was applied to the R^2 arm.

Utility values for health states PF and PP on and off treatment were estimated by means of a mixed effects model using EQ-5D-3L data collected in AUGMENT. As the disease characteristics that were used to derive utility values from the mixed effects model were population dependent, the utility values for R^2 versus R-CHOP/R-CVP and R^2 versus R-mono differed by population (see Table 1). The utility values resulting from the mixed effects model were used to inform the health states in the model for all treatments. Utility values from the study of Wild et al. [10], which were substantially lower for patients in particularly the PP state, were tested in a scenario analysis. Utility decrements for grade 3 and four adverse events were applied in the model for the expected duration of each adverse event, based on literature and previous appraisals.

The cost categories included in the model were costs associated with treatment (drug acquisition costs including subsequent therapies, drug administration costs including subsequent therapies, costs associated with treatment-related adverse events), disease monitoring costs, and costs associated with end of life care. All costs were based on or inflated to the 2018 price level. Unit prices were based on the NHS reference costs [11], Personal Social Services Research Unit (PSSRU) [12], Monthly Index of Medical Specialities (MIMS) [13], and the electronic Market Information Tool (eMIT) [14]. Dosing data for lenalidomide were taken directly from AUGMENT. Cost calculations were adjusted for treatment reductions and missed treatment cycles. The same method was applied to calculate rituximab costs for the R^2 arm. Drug administration costs were based on NHS reference costs tariffs, pharmacy costs for the preparation of the infusion, and NHS transport costs [11]. Costs of a full blood count were added to each treatment cycle for lenalidomide per visit to monitor the dose-limiting toxicities of neutropenia and thrombocytopenia. Costs of disease monitoring were separately estimated per health state and based on previous FL submissions [15, 16]. Costs of autologous stem-cell transplant (ASCT) were assigned to 11.8% of patients in R-CHOP. For R-CVP and R^2 , ASCT was considered not to occur in clinical practice and therefore there were no costs of ASCT in these comparators. The frequency of grade 3–4 adverse events that occurred in $\geq 2\%$ of patients was applied to the incidence rate for each treatment to obtain a one-off upfront cost for each treatment arm in the model. Terminal care was also applied as a one-off cost when a patient died. Lastly, subsequent treatments were applied in the model as an average one-off cost to patients entering the PP (on-treatment) health state, based on AUGMENT data for R^2 and the HMRN database for R-CHOP and R-CVP.

Fig. 1 Company’s model structure for treated follicular lymphoma and marginal zone lymphoma



In the company’s base-case analysis for the FL-only population, total life years and QALYs gained, as well as total costs, were higher in the R^2 arm compared with the R-CHOP and R-CVP arm. Incremental QALYs were mainly driven by QALY gains in the PP (off-treatment) health state. Incremental costs mainly resulted from higher drug acquisition costs. All cost and QALY results were confidential. The deterministic (probabilistic, based on 1000 iterations) incremental cost-effectiveness ratio (ICER) amounted to £15,909 (£27,768) per QALY gained for R^2 versus R-CHOP and £23,746 (£41,602) per QALY gained for R^2 versus R-CVP. For R^2 versus R-CHOP, the ICER was most sensitive to the cost of ASCT, the total subsequent treatment costs for R-CHOP and the proportion of patients who receive ASCT. For R^2 versus R-CVP, the ICER was most sensitive to the total subsequent treatment costs for R-CVP (including ASCT costs), administration costs, and resource use costs. The considerable difference between deterministic and probabilistic ICERs was attributed to increased uncertainty in the R^2 OS

extrapolations in the FL-only population compared to the initial FL + MZL population.

Similarly, for R^2 versus R-mono, the company’s base-case analysis (provided after the clarification phase upon request of the ERG) resulted in higher total life years and QALYs gained and higher costs for R^2 . Incremental QALYs were mainly driven by QALY gains in the PF health state. The cost difference was mainly caused by higher drug acquisition costs. The deterministic ICER amounted to £20,274 per QALY gained, and the probabilistic ICER was £23,412 per QALY gained. The deterministic sensitivity analysis revealed that the ICER was most sensitive to the total subsequent treatment costs for R^2 and R-mono and the frequency of haematologist visits PP.

3.4 Critique of Cost-Effectiveness Evidence and Interpretation

Searches were clear, transparent, and reproducible and unlikely to have missed any relevant studies. The ERG

Table 1 Health state utility values used in the economic FL-only model

Health state	R^2 vs. R-CHOP/R-CVP	R^2 vs. R-mono	Alternative values from Wild et al. [10]
Progression-free	0.867	0.846	0.805
Progressed (off treatment)	0.841	0.820	0.736 ^a
Progressed (on treatment)	0.806	0.785	0.620 ^b

R^2 lenalidomide in combination with rituximab, *R-CHOP* rituximab combined with cyclophosphamide, doxorubicin, vincristine, and prednisolone, *R-CVP* rituximab combined with cyclophosphamide, vincristine, and prednisolone, *R-mono* rituximab monotherapy

^aAssuming combined health states of active disease—newly diagnosed/relapsed

^bAssuming single health state active disease—relapsed

agreed with a de novo approach to modelling the cost-effectiveness of R^2 . The CS was largely in line with the NICE reference case, but deviated from the scope concerning the comparators modelled. More specifically, R-mono was excluded while direct evidence existed for R^2 versus R-mono, and in the refractory population, O-Benda was the sole comparator while NICE had explicitly stated it was not a relevant comparator for this appraisal. Most crucially, the ERG had concerns about the appropriateness of the PSM approach and its superiority over an STM and would have liked to see both approaches properly explored, particularly in the light of the limitations of PSM highlighted in NICE Technical Support Document (TSD) 19 [17]. PSM models have the advantage that they are easy to estimate from the trial time to event data, and because such data are employed to summarise treatment effectiveness, they are also easy to explain. However, they have a major disadvantage that each time to event function used to calculate the probability of remaining in each health state (PF or PD) is estimated independent of the other. Not only does this method lead to bias in that it is unlikely that the functions are not correlated, but it often leads to implausible scenarios such as the probability of remaining in the PF state exceeding the probability of remaining alive. Indeed, in this model, the curves were adjusted to ensure that long-term PFS estimates would not be higher than TTNLT or OS. Also, avoiding implausible curve crossing seemed to be the main argument for selection of survival function. Although the ERG requested the company provide an STM during the clarification phase, the company did not provide it until late in the process, and it only contained R^2 and R-mono as comparators, which hampered the ERG's assessment of the implications of using a PSM approach.

The ERG was concerned about the company pooling MZL and FL populations in the model, assuming they were comparable. The ICER for the company's FL-only scenario was substantially higher for the R-CHOP and R-CVP comparisons. This raises serious doubts about the validity of this assumption, and the ERG considered this to be a relevant source of uncertainty. In the re-submitted model following the final marketing authorisation that was granted for the FL population only, this was no longer an issue.

A main concern of the ERG was the trustworthiness of R^2 efficacy estimate resulting from the indirect comparison, which seemed to be inflated relative to the direct comparison data from AUGMENT. This could be concluded from the fact that QALYs for R^2 were substantially lower in the R^2 versus R-mono (direct) comparison than in the R^2 versus R-CHOP/R-CVP (indirect) comparison. So, the efficacy of R^2 was sensitive to the method used and therefore may have been biased. Although the ERG did not have the necessary data to quantify this uncertainty,

the use of efficacy estimates from the MAIC may have impacted the ICER substantially in favour of R^2 .

The ERG had concerns about the way survival curves were selected and validated. For the FL-only analyses presented in the company addendum, OS as predicted by the parametric survival curves was very different from OS curves presented in the original submission (which included both FL and MZL populations). No clinical validation of these new OS curves was performed. The ERG considered this process to deviate from TSD 14 recommendations [18] on survival analysis. The choice of OS likely introduced substantial uncertainty in the analyses.

The ERG considered utility values to be potentially overestimated, being higher than or comparable to those in the general population. With utilities remaining high throughout the model, any adjustment in survival curves had little impact on the ICER, as a high utility PP (relative to pre-progression) implied there was hardly any penalty on progression in terms of quality of life.

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

3.5 Additional Work Undertaken by the ERG

Based on all considerations highlighted in the ERG critique, the ERG defined a new base case for the FL-only population, in which various adjustments were made to the company's base case. This included correction of an operational error in the implementation of the "van Oers" scenario for R-CHOP efficacy, using subsequent treatment rates for R-CHOP and R-CVP taken from the pooled R-CHOP/R-CVP population instead of from a larger mixed R-chemo population, and capping utilities at the general population level. Furthermore, the ERG applied all six possible distributions to extrapolate OS in both arms. This was decided based on the divergent results of the different OS curves and the substantial uncertainty surrounding parametric survival model selection. In addition, exclusively for the R^2 versus R-CHOP and R-CVP comparisons, the log-logistic distribution was used to estimate PFS in the R^2 arm, and Weibull was used to estimate PFS in the R-CHOP/R-CVP arm. In this analysis, TTNLT was estimated with a log-logistic distribution in both arms. The probabilistic ERG base case for R^2 versus R-CHOP

ranged from £16,874 to £44,888 per QALY gained (based on 1000 iterations). For R^2 versus R-CVP, the ICER ranged from £23,135 to £59,810 per QALY gained, and for R^2 versus R-mono, it ranged from £18,779 to £27,156 per QALY gained.

Furthermore, the ERG explored alternative PFS distributions and treatment waning effects, an alternative source for adverse events in R-CHOP and R-CVP, the application of the same subsequent treatment costs for R^2 as for R-CHOP/R-CVP, lowered utilities, and an alternative source for R-CHOP efficacy. Applying the PP utility value by Pereira et al. [19] (0.45) was the most influential scenario (ICER R^2 vs. R-CHOP £33,626 per QALY gained, ICER R^2 vs. R-CVP £47,281 per QALY gained) that was explored by the ERG.

3.6 Conclusions of the ERG Report

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

Even though the ERG base-case ICER for R^2 versus R-CHOP was below £20,000 per QALY gained, the uncertainty around the cost-effectiveness of R^2 was substantial, mainly caused by the possible bias introduced by the indirect treatment comparison, which could not be accounted for in the ERG analyses. In addition, specific to the FL-only population analyses presented in the company addendum [20], the uncertainty around the OS estimates and the lack of clinical validation of these estimates would warrant even more caution in the interpretation of results. The ICER for R^2 versus R-CVP is higher and suffers from the same uncertainty.

4 Key Methodological Issues

The company chose to use a PSM. Because of compromises in choice of survival model resulting from implausible curve crossing, the ERG requested a scenario analysis using an STM during the clarification phase of the submission process. The ERG also requested an STM because TSD 19 [17] includes an explicit recommendation (number 11) saying

that “state transition modelling should be used alongside the PSM approach to assist in verifying the plausibility of the PSM extrapolations and to address uncertainties in the extrapolation period, even if this is only plausible for the pivotal trial”. The company did not provide an STM in their response. Their main argument to justify this was that survival data for the main comparators, R-CHOP and R-CVP, were not taken from a head-to-head trial with R^2 but from a registry. As this “real-world evidence” did not include regularly assessed disease progression status, the company considered it dubious to derive eventual OS estimates from intermediary events related to disease progression. Later on in the STA process the company did present an STM, but it was of very little value for cross-validation, because it did not include R-CHOP and R-CVP as comparators. In addition, there were different opinions at the committee meeting on whether the ERG should have asked the company for an STM, especially given the limited comparability between the two approaches in this particular case. This STA therefore illustrates how, on request of the ERG and in line with TSD 19, a cross-validation of PSM and TSM can be attempted. It also illustrates, despite the TSD 19 recommendations, that feasibility may be limited and how individual committees may have a different view. This may have to do with limitations of the STM approach as discussed above, or the added complexity of implementing both approaches, or other barriers to implementing the TSD 19 recommendation. The ERG, having experienced similar difficulty in a previous STA [21], therefore argues that care should be taken to justify the employment of this recommendation and feels that perhaps TSD 19 may need further elaboration to detail in what specific cases validation of PSM with STM is indicated. The ERG so far has not seen any STA having successfully cross-validated a PSM by an STM alongside it—or the other way around.

The original submission by the company included O-Benda as a comparator for rituximab refractory patients. However, NICE did not consider O-Benda a relevant comparator for disease that is refractory to rituximab, because O-Benda is only used as part of the Cancer Drugs Fund (CDF). This means that there is significant remaining clinical uncertainty, which needs more investigation through data collection in the NHS or clinical studies. The cost-effectiveness of drugs recommended for use within the CDF has not yet been established, and therefore any comparison of effectiveness or cost-effectiveness with CDF drugs are equally uncertain. It is therefore advisable that companies do not include comparators outside the scope in their submissions, as these will be ignored in the appraisal. On the other hand, there may be comparators that only become relevant after the final scope has been issued. As highlighted by Grimm et al. [22], it is important to also include the possibility of addition of comparators under appraisal at the time.

The company performed an MAIC: R^2 versus pooled data for R-CHOP/R-CVP for non-rituximab refractory patients using data from the HMRN. The use of MAICs remains largely untested; and there is a lack of clarity as to whether the results are relevant to the decision problem. The literature distinguishes between anchored and unanchored comparisons depending on whether a common comparator arm is used or not. Unanchored comparisons make much stronger assumptions and are widely regarded as infeasible [23].

The modelling of the treatment waning effect produced counter-intuitive results: assuming a later time point for treatment waning resulted in an increased ICER. This counter-intuitive result was most likely caused by the different shapes of the hazard functions, which are set to be equal when treatment waning kicks in. So, the ICER can be impacted substantially and in either direction by the choice of time point and the shape of the hazard functions. As the choice of the treatment waning starting point is usually highly uncertain, the ERG stresses the importance of checking the plausibility of any approach to extrapolating hazards over an extended period of time in any STA.

5 National Institute for Health and Care Excellence Guidance

On 7 April 2020, NICE recommended lenalidomide with rituximab, within its marketing authorisation, as an option for previously treated FL (grade 1– to 3A) in adults. It is only recommended if the company provides lenalidomide according to the commercial arrangement.

5.1 Consideration of Clinical Effectiveness

Clinical evidence for lenalidomide with rituximab and rituximab with chemotherapy is compared using an MAIC. R-CHOP and R-CVP are assumed to be clinically equivalent, although no evidence for this was presented by the company. The MAIC is as closely matched as possible, but relies on strong assumptions that are seldom met in reality.

5.2 Consideration of Cost-Effectiveness

The committee considered the PSM structure to be appropriate. The committee agreed that health-related quality-of-life values for lenalidomide with rituximab should be capped in the economic model to avoid having utility values higher than in the general population. The committee deemed that a

5-year treatment effect duration for lenalidomide with rituximab is appropriate, and that the exponential distribution is appropriate for extrapolating OS. The committee agreed that in extrapolating PFS, different distributions were needed for R^2 and R-CHOP/R-CVP. Finally, the committee concluded that given the most plausible range of ICERs, the combination of lenalidomide with rituximab can be considered a cost-effective use of NHS resources.

6 Conclusions

This article describes the STA considering lenalidomide in combination with rituximab for adults with previously treated FL or MZL. Following final marketing authorisation obtained for only FL, the STA focused on this population.

This STA illustrates the difficulty with the TSD 19 recommendation that ideally an STM should be provided alongside a PSM to verify the plausibility of extrapolations of the PSM. This recommendation is very rarely brought to practice, and even with an STM provided, as in this case, it was not straightforward to use it for verification purposes, in the absence of the relevant comparators.

Despite the uncertainty introduced by the use of an unanchored MAIC, which could not be accounted for in the economic modelling, and a few more concerns of the ERG, such as substantial uncertainty in the final OS extrapolations and potential overestimation of utility scores, the committee ruled that R^2 can be considered a cost-effective use of NHS resources. It therefore recommended R^2 as an option for previously treated FL in adults, when provided according to the commercial arrangement.

Acknowledgements This summary of the ERG report was compiled after NICE issued the FAD. All authors have commented on the submitted manuscript and have given their approval for the final version to be published. The views and opinions expressed therein are those of the authors and do not necessarily reflect those of NICE or the Department of Health. Any errors are the responsibility of the authors.

Declarations

Funding This project was funded by the National Institute for Health Research (NIHR) Health Technology Assessment (HTA) Programme. Please visit the HTA programme website for further project information (<https://www.nihr.ac.uk/funding-and-support/funding-for-research-studies/funding-programmes/health-technologyassessment>).

Conflict of interest WW, RR, NA, SR, SD, VHC, PP, GW, XP, BR, JK, MJ, SG, and AvA have no conflicts of interest to declare.

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

Availability of data and material Not applicable.

Code availability Not applicable.

Author contributions All authors have commented on the submitted manuscript and have given their approval for the final version to be published. RR, SD, and JK critiqued the clinical effectiveness data reported by the company. VHC and PP critiqued the literature searches undertaken by the company. GW critiqued the statistical analyses performed by the company. WW, NA, SR, XP, BR, MJ, SG, and AvA critiqued the mathematical model provided and the cost-effectiveness analyses submitted by the company. AvA acts as overall guarantor for this article. This article has not been externally peer reviewed by *PharmacoEconomics*.


Open Access This article is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License, which permits any non-commercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc/4.0/>.

References

- Riemsma R, Van Asselt ADI, Witlox W, Huertas Carrera V, Posadzki P, Armstrong N, et al. Lenalidomide with rituximab for previously treated follicular lymphoma and marginal zone lymphoma: a Single Technology Assessment. York: Kleijnen Systematic Reviews Ltd, 2019. <https://www.nice.org.uk/guidance/ta627/evidence>. Accessed 7 Jul 2020.
- National Institute for Health and Care Excellence. Lenalidomide with rituximab for previously treated follicular lymphoma and marginal zone lymphoma. Final scope. London: National Institute for Health and Care Excellence, 2019. <https://www.nice.org.uk/guidance/gid-ta10323/documents/final-scope>. Accessed 25 May 2019.
- Celgene Ltd. Lenalidomide for treated follicular lymphoma and marginal zone lymphoma [ID1374]. Response to request for clarification from the ERG. August 2019. Celgene Ltd, 2019.
- Leonard JP, Trneny M, Izutsu K, Fowler NH, Hong X, Zhu J, et al. AUGMENT: a phase III study of lenalidomide plus rituximab versus placebo plus rituximab in relapsed or refractory indolent lymphoma. *J Clin Oncol*. 2019;37(14):1188–99.
- van Oers MH, Klasa R, Marcus RE, Wolf M, Kimby E, Gascoyne RD, et al. Rituximab maintenance improves clinical outcome of relapsed/resistant follicular non-Hodgkin lymphoma in patients both with and without rituximab during induction: results of a prospective randomized phase 3 intergroup trial. *Blood*. 2006;108(10):3295–301.
- Sehn LH, Chua N, Mayer J, Dueck G, Trněný M, Bouabdallah K, et al. Obinutuzumab plus bendamustine versus bendamustine monotherapy in patients with rituximab-refractory indolent non-Hodgkin lymphoma (GADOLIN): a randomised, controlled, open-label, multicentre, phase 3 trial. *Lancet Oncol*. 2016;17(8):1081–93.
- Haematological Malignancy Research Network (HMRN). Clinical management and outcome of follicular lymphoma (FL) with a focus on relapsed/refractory disease (draft report). 17 July 2019. Haematological Malignancy Research Network, 2019.
- National Institute for Health and Care Excellence. Obinutuzumab with bendamustine for treating follicular lymphoma refractory to rituximab. NICE technology appraisal guidance 472. London: National Institute for Health and Care Excellence, 2017. <https://www.nice.org.uk/guidance/ta472>. Accessed 25 May 2019.
- National Institute for Health and Care Excellence. Rituximab for the treatment of relapsed or refractory stage III or IV follicular non-Hodgkin's lymphoma. NICE technology appraisal guidance 137. London: National Institute for Health and Care Excellence, 2008. <https://www.nice.org.uk/guidance/ta137>. Accessed 1 Feb 2019.
- Wild D, Walker M, Pettengell R, Lewis G. Pcn62 utility elicitation in patients with follicular lymphoma. *Value Health*. 2006;9(6):A294.
- NHS Improvement. NHS reference costs 2017/18. London: NHS Improvement, 2018. <https://improvement.nhs.uk/resources/reference-costs/>. Accessed 26 Apr 2019.
- Personal Social Services Research Unit. Unit costs of health and social care 2018. Canterbury: University of Kent, 2018. <https://www.pssru.ac.uk/project-pages/unit-costs/unit-costs-2018/>. Accessed 25 Jul 2019.
- MIMS. Monthly Index of Medical Specialities (MIMS) Online. 2020. Available from: <https://www.mims.co.uk/>. Accessed 7 Jul 2020.
- GOV.UK. Drugs and pharmaceutical electronic market information tool (eMIT). 2019. Available from: <https://www.gov.uk/government/publications/drugs-and-pharmaceutical-electronic-market-information-emit>. Accessed 25 Apr 2019.
- National Institute for Health and Care Excellence. Obinutuzumab for untreated advanced follicular lymphoma. NICE technology appraisal guidance 513. London: National Institute for Health and Care Excellence, 2018. <https://www.nice.org.uk/guidance/ta513>. Accessed 5 Jan 2019.
- National Institute for Health and Care Excellence. Rituximab for the first-line treatment of stage III-IV follicular lymphoma. NICE technology appraisal guidance 243. London: National Institute for Health and Care Excellence, 2012. <https://www.nice.org.uk/guidance/ta243>. Accessed 26 Apr 2019.
- Woods B, Sideris E, Palmer S, Latimer N, Soares M. NICE DSU technical support document 19: partitioned survival analysis for decision modelling in health care: a critical review. Sheffield: NICE Decision Support Unit, 2017. <http://www.nicedsu.org.uk>. Accessed 28 May 2019.
- Latimer N. NICE DSU technical support document 14: survival analysis for economic evaluations alongside clinical trials - extrapolation with patient-level data [Internet]. Sheffield: NICE Decision Support Unit, 2011. <http://www.nicedsu.org.uk>. Accessed 10 May 2019.
- Pereira C, Negreiro F, Silva C. PSY43 Economic analysis of rituximab in combination with cyclophosphamide, vincristine and prednisolone in the treatment of patients with advanced follicular lymphoma in Portugal. *Value Health* 2010;13(7):A468.
- Celgene Ltd. Lenalidomide with rituximab for treated follicular lymphoma and marginal zone lymphoma [ID1374]. Addendum for the amended follicular lymphoma only population. Company evidence submission to National Institute for Health and Care Excellence. Single Technology Appraisal (STA). Celgene Ltd, 2019.

21. Witlox WJA, van Asselt ADI, Wolff R, Armstrong N, Worthy G, Chalker A, et al. Durvalumab for the treatment of locally advanced, unresectable, stage III non-small cell lung cancer: an Evidence Review Group perspective of a NICE Single Technology Appraisal. *Pharmacoeconomics*. 2020;38(4):317–24.
22. Grimm SE, Fayter D, Ramaekers BLT, Petersohn S, Riemsma R, Armstrong N, et al. Pembrolizumab for treating relapsed or refractory classical Hodgkin lymphoma: an Evidence Review Group perspective of a NICE Single Technology Appraisal. *Pharmacoeconomics*. 2019;37(10):1195–207.
23. Phillippo DM, Ades AE, Dias S, Palmer S, Abrams KR, Welton NJ. Methods for population-adjusted indirect comparisons in health technology appraisal. *Med Decis Making*. 2018;38(2):200–11.

Affiliations

Willem J. A. Witlox¹ · Sabine E. Grimm¹ · Rob Riemsma² · Nigel Armstrong² · Steve Ryder² · Steven Duffy² · Vanesa Huertas Carrera² · Pawel Posadzki² · Gillian Worthy² · Xavier G. L. V. Pouwels^{1,4} · Bram L. T. Ramaekers¹ · Jos Kleijnen^{2,3} · Manuela A. Joore^{1,3} · Antoinette D. I. van Asselt^{1,5,6} 

¹ Department of Clinical Epidemiology and Medical Technology Assessment, Maastricht University Medical Centre+, Maastricht, The Netherlands

² Kleijnen Systematic Reviews Ltd, York, UK

³ Care and Public Health Research Institute (CAPHRI), Maastricht University, Maastricht, The Netherlands

⁴ Department of Health Technology and Services Research, Faculty of Behavioural, Management and Social Sciences, Technical Medical Centre, University of Twente, Enschede, The Netherlands

⁵ Department of Epidemiology, University Medical Center Groningen, University of Groningen, Hanzeplein 1, PO Box 30.001, 9700 RB Groningen, The Netherlands

⁶ Department of Health Sciences, University Medical Center Groningen, University of Groningen, Hanzeplein 1, PO Box 30.001, 9700 RB Groningen, The Netherlands