

Obinutuzumab in combination with chemotherapy for the first-line treatment of patients with advanced follicular lymphoma. An evidence review group evaluation of the NICE single technology appraisal

FW Thielen, NC Büyükkaramikli, R Riemsma, D Fayter, N Armstrong, C-Y Wei, V Hueartas Carrera, K Misso, G Worthy, J Kleijnen, I Corro Ramos

Pharmacoeconomics. 2019;37(8):975-984

ABSTRACT

Introduction: The National Institute for Health and Care Excellence (NICE), as part of the institute's single technology appraisal (STA) process, invited the company of obinutuzumab (Roche Products Limited) to submit evidence of the clinical and cost-effectiveness of the drug in combination with chemotherapy, with or without obinutuzumab as maintenance therapy for adult patients with untreated, advanced follicular lymphoma (FL) in the UK. Kleijnen Systematic Reviews Ltd (KSR), in collaboration with Erasmus University Rotterdam, was commissioned to act as the Evidence Review Group (ERG). This paper describes the company's submission, the ERG review, and NICE's subsequent decisions.

Methods: The clinical evidence was derived from two phase III, company-sponsored, randomised, open-label studies. Most evidence on obinutuzumab was based on the GALLIUM trial that compared obinutuzumab in combination with chemotherapy as induction followed by obinutuzumab maintenance monotherapy to rituximab in combination with chemotherapy as induction followed by rituximab maintenance monotherapy in previously untreated patients with FL (grades 1-3a). Long-term clinical evidence was based on the PRIMA trial, studying the benefit of two years of rituximab maintenance after first-line treatment in patients with FL. The cost-effectiveness evidence submitted by the company relied on a partitioned survival cost-utility model, implemented in Microsoft® Excel.

Results: The base case incremental cost-effectiveness ratio (ICER) presented in the company submission was below 20,000 GBP per quality-adjusted life-year (QALY) gained. Although the ERG concluded that the economic model met the NICE reference case to a reasonable extent, some errors were identified and several assumptions made by the company were challenged. A new base case scenario produced by the ERG suggested an ICER that was higher than the company base case, but still below £30,000 per QALY gained. However, some ERG scenario analyses were close to or even above the threshold. This was the case in particular for assuming a treatment effect that did not extend beyond trial follow-up.

Discussion: These results lead to an initial negative recommendation by the appraisal committee. Subsequently, the company submitted a revised base case focusing on patients at intermediate or high risk of premature mortality. Simultaneously, a further price discount for obinutuzumab was granted. In addition to the company's revised base case, the ERG suggested a restriction of the treatment effect to five years and implemented biosimilar uptake and cheaper prices for rituximab. All of these adjustments did not exceed £30,000 per QALY gained and therefore the use of obinutuzumab for patients with advanced FL and a FLIPI score of two or more could be recommended.

INTRODUCTION

Within the United Kingdom (UK) National Health Service (NHS), the National Institute for Health and Care Excellence (NICE) may recommend the use of new and existing medicines and treatments through single technology appraisals (STAs).³²² During such appraisals, the NICE Appraisal Committee (AC) reviews clinical and economic evidence on the technology under investigation based on the company submission (CS) taking into account the critique of a report from an appointed independent Evidence Review Group (ERG) as well as advice from other consultees (e.g. patients, experts and other stakeholders). After consideration of all the relevant evidence, the AC formulates preliminary guidance in the form of the Appraisal Consultation Document (ACD) as to whether or not to recommend the technology. The stakeholders are invited to comment on this ACD and the submitted evidence. A subsequent ACD may be produced or a Final Appraisal Determination (FAD) is issued. Once recommended by NICE, the NHS is legally obliged to fund and resource the appraised technology so that the patients' right to access these technologies is ensured.³²² This paper presents a summary of the ERG report and the development of the NICE guidance based on the AC's findings for the STA of obinutuzumab in combination with chemotherapy as an induction therapy followed by obinutuzumab maintenance monotherapy, for the first-line treatment of patients with advanced follicular lymphoma. Full details of all the relevant appraisal documents can be found on the NICE website (TA513).³²³

THE DECISION PROBLEM

Arising from lymphocytes, lymphomas are a heterogeneous group of malignancies of which about 90% are diagnosed as non-Hodgkin lymphomas (NHL).³²⁴ In the UK, NHL are the sixth most common type of cancer.³²⁵ With an annual incidence rate of 3.2 per 100,000, follicular lymphomas (FL) constitute the third most common subtype of NHL in UK.³²⁵ The median age at diagnosis for FL in the UK is about 65 years with a five-year relative survival rate of 86.5%.^{326,327} Survival of patients with FL can be predicted with either the Follicular Lymphoma International Predictive Index (FLIPI),³²⁸ or its revised version, known as FLIPI2.³²⁹ The FLIPI comprises a set of five predictive parameters (age, Ann Arbor stage, haemoglobin, serum LDH level, and number of nodal sites), discriminating patients into three risk groups (low, intermediate and high).³²⁸ FLIPI2 builds on five parameters as well, but only haemoglobin and age are shared with the FLIPI (β 2-microglobulin, longest diameter of the largest involved node, and bone marrow involvement are the other three).³²⁹ The lower the FLIPI/FLIPI2 score, the better the prognosis as well as overall survival (for FLIPI) and progression free survival (for FLIPI2) of patients with FL. A FLIPI score of 0 to 1 is regarded as low risk while a score of 2 or above signifies intermediate to high risk.

Due to the indolent nature of the disease, most patients present with advanced stages at first diagnosis and therefore require systemic treatment.^{326,330}

In the UK, patients with advanced, symptomatic FL, receive rituximab-containing treatment regimens as first-line therapy options, followed by two years maintenance therapy. However, despite receiving immunochemotherapy, an estimated 20% of patients with FL experience disease progression within two years from diagnosis.³³¹ Treatment for patients with advanced stage targets to control the disease, which is typified by a chronic course comprised of repeated relapses, treatment and progression.

As a Type II anti-CD20 antibody, obinutuzumab received marketing authorisation from the European Medicine Agency (EMA) for previously untreated chronic lymphocytic leukaemia (CLL) in July 2014.³³² In April 2016 and July 2017, the EMA extended the indication for obinutuzumab to FL.^{333,334} The scope for this STA set by NICE was to assess the use of obinutuzumab in combination with chemotherapy, with or without obinutuzumab as maintenance therapy (obin-chemo+obin) for people with untreated advanced FL in the UK. Mentioned comparators were rituximab monotherapy (although there was no marketing authorisation in the UK for this indication at time of appraisal), rituximab-based chemotherapy, with or without rituximab maintenance treatment, and bendamustine monotherapy (no marketing authorisation in the UK at the time of appraisal).

THE INDEPENDENT EVIDENCE REVIEW GROUP (ERG) REVIEW

Kleijnen Systematic Reviews Ltd (KSR), in collaboration with the Erasmus University Rotterdam constituted as the ERG and reviewed the evidence on the clinical and cost-effectiveness of obin-chemo+obin treatment for the first-line treatment of patients with advanced follicular lymphoma.

The ERG critically reviewed the evidence in the CS, the company's responses to clarification questions (RCQ) from the ERG, and the evidence provided after the publication of the ACD. Furthermore, the ERG explored the impact of assumptions on the incremental cost-effectiveness ratio (ICER), revised the economic model and explored additional scenario analyses.

Summary of the clinical evidence

The company conducted a systematic review to identify published and unpublished randomised clinical trial (RCT) evidence on the use obinutuzumab in previously untreated FL. Searches were carried out in June 2015 via PubMed, Embase and the Cochrane Central Register of Controlled Trials (CENTRAL). Update searches were reported for March and June 2017. At each time, separate supplementary searches of congress proceedings, clinical

trial registries, cancer association networks, and HTA agency websites were conducted to detect relevant unpublished and grey information. Ultimately, only one study was considered relevant to the decision problem. Therefore, the evidence on the efficacy of obinutuzumab was entirely based on GALLIUM, a phase III, open-label, multicentre RCT (BO21223, NCT01332968).³³⁵

GALLIUM was conducted at 177 trial centres in 18 countries with 21% of the participants (n=293) being from the UK. For this study, 1401 adult patients (≥18 years) with previously untreated, advanced iNHL were randomly assigned to two treatment arms. All analyses were based on the 1202 patients in the FL subpopulation, which included those with FLIPI scores from 0 upwards. One group received rituximab in combination with chemotherapy (R-chemo) as induction followed by rituximab (R) as maintenance. The other group received obinutuzumab in combination with chemotherapy (obin-chemo) as induction, followed by obinutuzumab (obin) as maintenance.

The primary outcome was investigator-assessed progression-free survival (INV-PFS). Secondary outcome measures included independent review committee-rated progression-free survival (IRC-PFS) and overall survival (IRC-OS), response rates at induction, maintenance and follow-up, adverse events (AEs), as well as health-related quality of life (HRQoL). The latter was assessed using the disease-specific Functional Assessment of Cancer Therapy for Patients with Lymphoma (FACT-Lym) instrument and the European Quality of Life (Euro-Qol) ED-5D-3L questionnaire.

Data reported in the CS were taken from the primary analysis of GALLIUM with a clinical cut-off in January 2016. Since the updated cut-off data were marked as academic in confidence, all data presented in this publication focus on the FL subpopulation at the January 2016 cut-off of the GALLIUM trial only.

The median age of the FL population in GALLIUM was 59 years (range: 23-88 years) with a female ratio of 53.2%. Of the three different chemotherapy regimens permitted, the most frequently used was bendamustine (57%), followed by CHOP (33%) and CVP (10%). Induction therapy was completed for 598 and 594 patients in the R-chemo and obin-chemo arm respectively. Subsequently, 527 patients in the R-chemo and 539 patients in the obin-chemo arm received maintenance therapy. These patients had achieved either a complete response (CR) or a partial response (PR) at this stage. Stable disease (SD) was achieved by 17 patients who therefore underwent observation. Patients in both maintenance and observation were followed clinically on a two-monthly basis for two years. A total of 341 patients in the R-chemo arm and 361 patients in the obin-chemo arm completed maintenance therapy. The reported overall median observation time was 34.4 months (range: 0.1–54.5 months) in the R-chemo arm and 34.8 months (range: 0.0–53.8 months) in the obin-chemo arm. The proportion of patients who had been observed for at least 2 and 3 years at the clinical cut-off was 87.7% and 44.1% in the R-chemo arm and 91.3% and 45.1% in the obin-chemo arm, respectively.

At the January 2016 cut-off, median INV-PFS as well as OS were not reached in both treatment arms. While the IRC-PFS was also not reached for the obin-chemo arm, it was reported with 51.2 (95% CI: 47.1 – not estimated) months for the R-chemo arm. Hazard ratios (HRs) comparing obin-chemo+obin to R-chemo+R were consistent and in favour for obin-chemo+obin in both INV-PFS (HR 0.66 [95% CI: 0.51, 0.85; $p = 0.0012$]) and IRC-PFS (0.71 [95% CI: 0.54, 0.93; $p = 0.0138$]). In the subgroup analyses, only low risk FLIPI scores showed a HR greater than 1.00 (HR 1.17 [95% CI: 0.63, 2.19]), while HRs of all other FLIPI scores ranged from 0.40-0.86.

Treatment-related AEs were observed in 94.8% of patients in the obin-chemo+obin arm and in 91.6% of patients in the R-chemo+R arm. Two System Organ Classes were more commonly observed ($\geq 10\%$) in the obin-chemo+obin arm when compared to the R-chemo+R: general disorders and administration site conditions (50.8% vs. 60.8%) as well as injury, poisoning and procedural complications (49.1% vs. 59.2%). In addition, a higher incidence of severe adverse event (SAEs) could be observed in the obin-chemo+obin arm than in the R-chemo+R arm (46.1% vs. 39.9%).

For both FACT-Lym questionnaire subscales and EQ-5D-3L scales, no notable differences between the treatment arms was detected at any time.

It is worthwhile mentioning that data for the late progressive disease states were taken from the PRIMA study, a phase III RCT of rituximab maintenance therapy in patients with high tumour burden FL that responded to rituximab plus chemotherapy induction. Characteristics of the PRIMA trial are reported elsewhere.^{336,337}

Critique and conclusion of the clinical evidence and interpretation

Regarding the systematic review, the ERG concluded that the population of the review was in line with the scope but the comparators were not. This was because the company had only included studies with a rituximab arm. As a deviation from the scope set by NICE, the company chose to consider only one comparator: rituximab in combination with chemotherapy, followed by rituximab maintenance treatment in patients achieving a response, which was criticised by the ERG. In addition, the ERG criticised that both data extraction and quality assessment of the studies retrieved by the literature search were not conducted by two reviewers independently. However, the ERG acknowledged that the GALLIUM trial was appropriate for the decision problem at hand due to its relevant population and reasonable proportion of UK patients (21%). The attention of the AC was drawn to the exclusion of grade 3B lymphoma in GALLIUM, which was found to be in line with the anticipated treatment with obinutuzumab. Although three different types of chemotherapy were offered in the trial, GALLIUM was not designed to investigate differences between these regimens. Therefore, the ERG could not decide whether the breakdown of regimens in the trial would reflect UK clinical practice. Due to GALLIUM being an open-label trial, the ERG concluded that the results of the IRC would be less prone to bias than the investigator

results. Furthermore, although the follow-up duration of the trial was reasonable, data were judged not fully mature for the main outcomes.

In its clarification response the company acknowledged that the GALLIUM cohort might on average be younger than the average UK patient population with FL which was confirmed by clinical experts consulted by the company and data of the Haematological Malignancy Research Network (HMRN)³³⁸ The ERG therefore adjusted the age at baseline in its own base case model.

Summary of the cost-effectiveness evidence

The company conducted several literature searches to detect studies on cost-effectiveness, health effects, as well as on cost and healthcare resource use. For the cost-effectiveness search, none of the retrieved references were considered relevant. Therefore, the company conducted a *de novo* economic evaluation of obinutuzumab in combination with chemotherapy compared with rituximab in combination with chemotherapy for the first-line treatment of patients with advanced FL.

For this purpose, the company developed an Excel-based five-state cohort transition Markov model with following health states: two progression-free survival (PFS) states for on and off treatment, two progressed disease states (PD) for early and late PD, and death. All simulated observations start in the PFS health state on treatment (obin-chemo+obin or R-chemo+R). Only patients responding to induction received maintenance (as per licence indication), which was only offered until progression or for a maximum of two years. Patients are considered 'off-treatment', when they complete or discontinue treatment in the PFS state. From the latter state, patients can remain in either PFS (on- or off-treatment) or transition due to a progressive disease or death. It is assumed that the time to progression after initial treatment is predictive for post progression mortality and overall survival. In particular, patients progressing within the first two years of initial treatment have significantly worse survival outcomes than those who did not progress that early.^{331,339} To account for different outcomes and costs to the cohorts of patients who experience an early or a late progression, two progressed disease (PD) states were introduced. Once patients enter any of the two PD states, they can only remain in this state until death. The cumulative deaths from PFS and early as well as late PD states were used to calculate overall survival in the model. Survival estimates beyond the observed trial duration were extrapolated using several parametric functions. The population considered in the *de novo* analysis was equal to the GALLIUM cohort in terms of average age, body weight, height und Body Surface Area (BSA) (see Table 1).

The company argued that, based on the observed long-term follow-up in the PRIMA study and the expert opinion from clinical advisors, there was no evidence of a finite duration of treatment effect in treatments of FL (including obin-chemo+obin and R-chemo+R). However, for the base case analysis a treatment effect duration of 9.75 years was assumed,

Table 1 – Patient baseline characteristics used in the company’s de novo economic evaluation

| Patient characteristic | Baseline value | Source |
|------------------------------------------------|----------------|-----------------------------------|
| Age (years) | 57.9 | GALLIUM trial data ³³⁵ |
| Body weight (kg) | 75.7 | |
| Height (cm) | 168.3 | |
| Calculated Body Surface Area (m ²) | 1.86 | |

based on the PRIMA study. This assumption was tested in the sensitivity analysis for its robustness with a minimum and maximum assumed treatment effect duration of five years and an infinite time duration, respectively.

The transition probabilities between the different model states were estimated from parametric survival functions fitted to the relevant data for the time to treatment discontinuation (TTTD), PFS, and post progression survival (PPS). Considered distribution were: exponential, Weibull, log-logistic, log-normal, gamma, and Gompertz. Their goodness of fit to the GALLIUM data was assessed with both the Akaike/Bayesian Information criterion (AIC/BIC).

In the model, OS was calculated as the sum of the time spent in both PFS and the two PD health states (early and late). For the PFS and early PD mortality estimates, the company relied on the investigator assessed (INV) data of the GALLIUM trial. Since in GALLIUM no PPS events in late progression were observed, late PD mortality estimates were based on the PRIMA study.

For the CS base case model, long-term PFS and early PD were modelled with an exponential distribution. The monthly probability of transitioning from PFS to death was based on the UK age-specific all-cause mortality rate or the PFS death rates in GALLIUM (whichever occurred first). The same method was applied for the post progression survival probabilities; except that late PD mortality rates were estimated using data from the PRIMA trial (instead of GALLIUM). This was necessary because the GALLIUM data was premature.

HRQoL utilities for the PFS states were based on EQ-5D values collected in the GALLIUM trial. However, since long-term EQ-5D values from this trial were considered immature, the utility inputs for the progressed model states were based on utilities elicited in another study that had originally been commissioned by Roche.³⁴⁰ Health state related costs used in the model consisted of costs for medication (induction and maintenance), supportive care, subsequent treatment in PD, transportation, and adverse events. Relevant medication costs included those of obinutuzumab, bendamustine, CHOP, CVP, and rituximab. The latter were based on 2017 UK reference prices.^{341,342} No vial sharing was assumed.

For both costs and utilities, a 3.5% discount rate was applied. In addition, the company conducted a one-way deterministic sensitivity analysis, several scenario analyses and a probabilistic sensitivity analysis (PSA) to explore parametric and structural uncertainty.

In the RCQ, the initial 40-year time horizon of the model was set to 50 years, following the ERG's request.

In the final model submitted by the company, the base-case ICER (cost per quality-adjusted life-year [QALY] gained) was lower than 20,000 GBP, the threshold set by NICE being somewhere between 20,000 GBP and 30,000 GBP, depending on plausibility of the evidence.³⁴³ In addition, in the PSA results, it was observed that the ICER never reached this threshold. Scenario analyses revealed that different assumptions on the length of treatment effect (no finite duration versus five years), discount values (3.5% vs. 1.5%), and the choice of the parametric survival curve (exponential vs. Weibull) for the PFS states had with the highest impact of the ICER but did not exceed the NICE threshold. In general, all ICERs presented by the company (base case, PSA, and scenario) were below 30,000 GBP per QALY gained.

Critique of the Cost-Effectiveness Evidence and Interpretation

After assessing the company's submitted evidence and model, the ERG concluded that, although the cost-effectiveness searches were well documented and reproducible, there were concerns regarding inclusion of a Line of Treatment facet that was overly restrictive. Revised searches were provided during the clarification process, which retrieved additional references. Despite the revision of the strategies, errors in the Cochrane Library search syntax were still present. The additional studies identified were not found to be relevant for the decision problem and were not included.

The economic model met the NICE reference case to a reasonable extent. Nevertheless, the ERG found that deviations might have occurred in both measurement and valuation of HRQoL. Other deviations included the company's choice of the intervention and the comparator. In general, the model was in line with the company's formulated decision problem but only partially in line with the scope. While the intervention in the scope was described as "*obinutuzumab in combination with chemotherapy, with or without obinutuzumab maintenance therapy*",³⁴⁴ the company assessed the obinutuzumab induction therapy with obinutuzumab maintenance therapy only. Hence, the cost-effectiveness of obinutuzumab in combination with chemotherapy induction without maintenance therapy was not explored in the submission. In terms of comparators, the company focussed on rituximab in combination with chemotherapy but did not include other relevant comparators listed in the NICE scope (such as rituximab mono-therapy or bendamustine mono-therapy).

For the model structure, the ERG concluded that the one used in the CS was slightly different from other, commonly used partitioned survival models in oncology.^{345,346} The transitions between the health states were explicitly modelled, and in addition to other models, the company had also incorporated early and late PD states, which was seen by the ERG as a valid addition for modelling FL patients. With regards to the analysis and extrapolation of the survival data, the company had followed the guidance from the NICE Decision Support

Unit (DSU).³⁰⁶ Nevertheless, the ERG criticised the choice of the exponential distribution for the PFS survival probabilities. Although the company had argued that the exponential distribution would result in a more conservative estimate when compared to the log-logistics distribution, the ERG considered the Weibull distribution would represent an even better estimate with similar AIC and BIC values. This was because the Weibull distribution predicted a 10-year PFS probability of 30.2%, which was in line with clinical expert opinion suggesting that approximately 60-70% of patients relapse in the first 10 years after treatment.

Furthermore, due to few mortality events in the GALLIUM study, the company had based the transition probabilities from PFS to death on pooled estimates between the treatment arms. Consequently, the probability of dying in PFS was assumed equal for both treatment arms. Yet, the number of observed deaths between the treatment arms differed and showed higher mortality in the obinutuzumab arm (although not statistically significant). Therefore, the ERG recommended applying different transition probabilities per arm for both PFS and post progression survival (early and late PD).

Another point of criticism by the ERG was the assumption of a finite duration of the treatment effect on PFS, which was the main driver of the cost effectiveness results. This assumption was made solely on data from the PRIMA trial, as long-term data from GALLIUM was lacking. In the model base case of the CS, a 9.75-year treatment effect was assumed although the longest follow-up in the GALLIUM trial (at the time of the submission) was five years. Therefore, the ERG considered a finite treatment effect of five years as a more conservative approach to model the cost effectiveness.

Similar to the critique of the clinical evidence, the ERG criticised the used on INV-PFS data instead of the IRC-PFS data.

Additional Exploratory Analysis Conducted by the ERG

After careful consideration of all input parameter assumptions in the company's base case, the ERG defined a new base-case scenario, including multiple adjustments to the company's base-case economic model. The adjustments were categorised following the suggestions of Kaltenthaler et al.³⁴⁷:

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

Ultimately, the ERG corrected five errors, three violations, had four matters of judgement, and tested four alternative scenarios (see online Appendices 7A and 7B for a complete list of all items). The most influential adjustments to the company's base case model were: (1) choosing the IRC-PFS data together with a Weibull distribution for the PFS extrapola-

tion, (2) applying age dependant utility decrements, (3) increasing the population's age at baseline, and (4) considering different mortality rates per treatment arm.

The ERG revised base-case ICER did not exceed the 30,000 GBP threshold per QALY gained. Simultaneously, in more than 50% of the PSA results, obin-chemo+obin was cost-effective when compared to R-chemo+R at a 30,000 GBP per QALY-gained threshold. For all but one scenario analyses, the estimated ICER remained below the 30,000 GBP threshold per QALY. Only assuming a treatment effect duration of five years (scenario 1a) yielded an ICER above 30,000 GBP per QALY gained. Choosing different sources for utilities in PFS and PD had a substantial impact on the ICER but did not exceed the mentioned threshold.

Conclusion of the ERG Report

The ERG concluded that the GALLIUM trial is a good quality RCT even though a number of limitations were found. For instance, the breakdown of chemotherapy regimens (CHOP, CVP, and bendamustine) received in the trial may not reflect UK clinical practise. Likewise, the median age of included individuals in GALLIUM was not reflective of the UK FL population. In addition, although the trial had a reasonable follow-up, data were not fully mature for the main outcomes. Consequently, the ERG expressed major concerns regarding the implementation of the treatment effectiveness. More specifically, choosing a shorter duration of the treatment effect than in the CS was a major driver for bringing the ICER close, or even above the 30,000 GBP per QALY-gained threshold. Choosing either INV-PFS or IRC-PFS data had a substantial impact on the ICER as well, although none of two scenarios exceeded the mentioned threshold. Other remaining concerns were related to PFS probability distributions, and choosing the same mortality rate for both treatment arms. In addition, the ERG could not verify the source of the stated utility values for the PD states since they were referenced with an abstract that did not present any EQ-5D values.³⁴⁰

Nevertheless, the ERG preferred base-case analysis resulted in an ICER lower than the 30,000 GBP per QALY threshold. Although the ICER seemed to be robust to most structural changes explored by the ERG, the scenario analyses conducted by the ERG showed that with different choices for the treatment effect duration or the utilities for the PD health state, the ICER could exceed the 30,000 GBP per QALY threshold.

ERG research recommendations

Most notable, the cost-effectiveness analyses could benefit from long-term follow-up results from GALLIUM. The results could validate some of the key assumptions made in the model such as the extrapolation of the PFS parametric survival functions, the duration of the treatment effect and the early/late mortality in the progressed disease state. Different mortality assumptions in the progressed disease state would lead to different PFS/OS surrogacy implications, which might have substantial impacts on incremental results as elaborated in other first-line treatment submissions in oncology.³⁴⁸ In addition, a more recent and transparent

measure of utility values for the PD health state would increase both validity and reliability of the HRQoL estimates. Future research should include a comparison of obinutuzumab with different chemotherapy regimens such as CHOP, CVP, and bendamustine to generate more reliable estimates for direct treatment comparisons.

Key methodological issues

Three issues were expressed by the ERG.

First, the company's assumption of a finite duration of treatment effect on PFS was a major driver for the ICER. The treatment effect duration was implemented by setting the hazard ratio of the modelled intervention to the comparator to one (i.e. no difference in treatment effect), once the end of the treatment effect was assumed. Although, the ERG deemed the technical realisation of the limited treatment effect duration appropriate, some methodical concerns on its duration were expressed. The company had based its assumption on the PRIMA study, where no finite duration of treatment effects between rituximab maintenance treatment compared with 'observation only' could be observed until the longest follow-up period of 9.75 years. Likewise, the assumption of proportional hazards between intervention and comparator arm seemed to hold throughout the PRIMA study. This was backed by clinical advisors who had suggested that there is no evidence of finite treatment effect in FL treatments, and that this might hold true for the comparison of obin-chemo+obin versus R-chemo+R. The ERG by contrast doubted the generalisability from the PRIMA results and their transferability to GALLIUM. A visual inspection also suggested that the log-cumulative hazard plots for PFS from GALLIUM converged. Hence, the proportional hazards assumption most likely did not hold. Based on the available evidence, no robust estimate alternative for a treatment effect duration could be given. On these grounds, the ERG proposed a duration of five years for the treatment effect, as this also reflected the longest follow-up time of the GALLIUM study.

Second, due to the immaturity on the GALLIUM data, the company used PRIMA data to model the late PD health states. This combination of evidence was done without any kind of adjustments. However, since patient characteristics between the two studies might not be comparable, an unadjusted use of these data might bias the model estimates.

Third, the company's assumption of no biosimilar uptake for rituximab was deemed implausible.

NICE GUIDANCE

Key issues Considered by the Appraisal Committee

Regarding the clinical evidence, the AC concluded that the population of the GALLIUM trial would reflect people with advanced FL receiving treatment within the NHS to a reason-

able extent. However, the trial was judged to be underpowered to show a difference in OS. In addition, the presented trial data was considered highly immature. Consequently, the AC could not conclude whether obinutuzumab could prolong overall survival when compared to rituximab. Although it was acknowledged that obinutuzumab delays disease progression in the short term, there was still uncertainty about a long-term effect on progression-free survival. Furthermore, the AC concluded that obinutuzumab is associated with higher burden of adverse events when compared to rituximab. In terms of HRQoL, the AC regarded the difference in the EQ-5D scores between the GALLIUM trial arms as not statistically significant.

For the cost-effectiveness analysis, the AC considered several key issues.

First, it recognised substantial uncertainty in the evidence base and ICER in terms of the treatment effect duration, extrapolation of progression free survival, resource use and the immaturity of the clinical data. In addition, the company's assumption on the treatment effect duration was criticised.

Second, the AC doubted the company's assumption of a 0% uptake of biosimilars for rituximab and was aware of two biosimilar versions of rituximab that had a market authorisation. In addition, the AC was informed that the current uptake was increasing and at around 40%.

Due to the substantial uncertainty in the evidence base and ICER, the AC concluded that an acceptable ICER threshold would not lie towards the upper part of the 20,000 GBP to 30,000 GBP per QALY gained range specified in the NICE's guide to the methods of technology appraisal.³⁴³ Other factors that could substantiate a 30,000 GBP per QALY gained such as the innovative nature of obinutuzumab or special consideration as a 'life-extending treatment at the end of life' were ruled out as well.

Ultimately, the ERG updated its base case analysis incorporating the ACs findings. This yielded an expected increase in the ICER, rendering the intervention not cost-effective at the threshold of 30,000 GBP per QALY gained.

Preliminary Guidance (First Appraisal Consultation Document [ACD])

For the first ACD, NICE had considered the initial evidence submitted by the company, the testimony of professional groups and other stakeholders, as well as the ERG report. In September 2017, NICE's first ACD did not recommend obinutuzumab within its marketing authorisation for untreated advanced FL in adults. This recommendation was however not intended to affect patients that had already started treatment with obinutuzumab before the guidance was published. For these patients no change in funding arrangements would take place. A second appraisal committee meeting was planned for October 2017. Until then, all consultees had the possibility to react to the previously presented evidence.

Response to Preliminary Guidance (First ACD)

In reaction to the first ACD, the company provided an updated version of its model including alternative assumptions and a revised base case population. The company argued that, although the trial was not powered for individual FLIPI subgroups, an analysis focussing only on intermediate and high FLIPI subgroups would result in sufficient PFS events for the modelling task. This was in line with the EMA requirement to include a statement in the obinutuzumab Summary of Product Characteristics that the “[...] efficacy in FLIPI low risk (0-1) patients is currently inconclusive [...]”,³⁴⁹

With regard to the duration of treatment effect, the company retained its earlier argumentation and assumed an infinite duration of treatment effect for obinutuzumab. Furthermore, the new model assumed independent PFS extrapolations for high and intermediate FLIPI subgroups (thus using a non-proportional hazards assumption) and implemented vial sharing.

Biosimilar uptake for rituximab was not considered for the base case but in scenario analyses. The company had based this assumption on the recent availability of biosimilar rituximab, claiming that the branded product would currently constitute the majority of IV rituximab used.

This new analysis yielded an ICER below £20,000 per QALY gained, whereas scenarios considering different proportions of market shares and price reductions for biosimilar rituximab showed an ICER above £20,000 per QALY.

In reaction to the new company model, the ERG proposed a new (final) base case that assumed a treatment effect duration of five years, independent PFS extrapolations for high and intermediate FLIPI subgroups (thus using a non-proportional hazards assumption), no vial sharing, and a 65% update of biosimilar IV rituximab.

In the meanwhile, the company had also agreed on a new patient access scheme that would provide a further discount to the list price of obinutuzumab. The level of the discount is, however, commercial in confidence. The new ERG base case (including the new patient access scheme) yielded an ICER below £30,000 per QALY gained.

Final Appraisal Determination (FAD)

Due to the revised economic analyses focussing on higher-risk subgroups and a further discounted price for obinutuzumab and rituximab, the ICER was estimated to be below £30,000 per QALY gained. Hence, the AC issued new guidance. In March 2018, the FAD recommended obinutuzumab as an option for untreated advanced FL in adults, restricted to patients with a FLIPI score of 2 or more (as intended by the company’s revised model), provided that the company would grant the negotiated simple price discount in the revised patient access scheme. Just as in the ACD, the recommendation was not intended to affect patients that had already started treatment with obinutuzumab before the guidance was published. For these patients no change in funding arrangements would take place.

CONCLUSIONS

This STA demonstrates that even an ICER of below £20,000 per QALY gained, as in the first company submission, is no guarantee for a positive recommendation by NICE. In fact, a high degree of uncertainty in the evidence base of the cost-effectiveness model might lead to a negative decision by NICE, even when the ERG's base case is below the £30,000 per QALY gained threshold. Instead, NICE is also valuing the degree of uncertainty that, in this case, was expressed by some initial ERG scenarios yielding an ICER close to and above the £30,000 threshold.

For this submission, the AC's major concerns were on the degree of uncertainty of the various cost-effectiveness model input parameters, particularly the treatment effect duration. Although in this STA most of these parameters were based on an RCT, the immaturity of its results did not allow for robust estimates concerning the long-term effects of obinutuzumab on progression-free survival. Furthermore, the AC acknowledged both cheaper prices and higher uptake of rituximab biosimilars. This was based on the 2017-2019 prescribed services commissioning for quality and innovation (CQUIN), issued by NHS England, encouraging the use of biosimilar products to reduce the costs for medicines.³⁵⁰ Consequently, the AC had assumed that most commissioners in England would prefer to purchase cheaper biosimilar versions rather than branded drugs.

Narrowing down the treatment indication to patients with a FLIPI score of 2 or higher, as well as providing a further discount on the price for obinutuzumab could reduce the degree of decision uncertainty to the extent that the AC could issue a positive recommendation. This final decision deviates from the originally intended patient group proposed by the company (no FLIPI score thresholds). Hence, when considering the scope of the first CS, NICE adopted a 'restricted' or 'optimized' decision to provide access to obinutuzumab while concomitantly reducing the decision uncertainty.^{351,352}