LIVERPOOL REVIEWS AND IMPLEMENTATION GROUP (LRiG)

Rituximab for the first-line maintenance treatment of follicular non-Hodgkin's lymphoma

This report was commissioned by the NIHR HTA Programme as project number 08/218

Completed 7th October 2010

UNIVERSITY OF LIVERPOOL

LIVERPOOL REVIEWS AND IMPLEMENTATION GROUP

A MEMBER OF THE RUSSELL GROUP

Rituximab for the first-line maintenance treatment of follicular non-Hodgkin's lymphoma

Produced by: Liverpool Reviews and Implementation Group (LRiG)

Authors: Adrian Bagust, Professor of Modelling in Health, Liverpool Reviews and Implementation Group, University of Liverpool

Angela Boland, Associate Director, Liverpool Reviews and Implementation Group, University of Liverpool

Michaela Blundell, Research Fellow (Medical Statistician) Liverpool Reviews and Implementation Group, University of Liverpool

Patrick Chu, Professor and Consultant Haematologist, Royal Liverpool Hospital, Liverpool

Yenal Dundar, Research Fellow, Liverpool Reviews and Implementation Group, University of Liverpool

Janette Greenhalgh, Research Fellow (Clinical Effectiveness), Liverpool Reviews and Implementation Group, University of Liverpool

Juliet Hockenhull, Research Fellow (Clinical Effectiveness) Liverpool Reviews and Implementation Group, University of Liverpool

James Oyee, Research Fellow (Medical Statistician), Liverpool Reviews and Implementation Group, University of Liverpool

Chris Proudlove, Director, North West Medicines Information Centre, Pharmacy Practice Unit, Liverpool

Correspondence to: Dr Janette Greenhalgh, Research Fellow, Liverpool Reviews and Implementation Group, University of Liverpool, Room 2.2, Whelan Building, The Quadrangle, Brownlow Hill, Liverpool L69 3GB

Date completed: 7th October 2010

Source of funding: This report was commissioned by the NIHR HTA Programme as project number 08/218

Declared competing interests of the authors: None

Acknowledgements:

Thanks to Ruben Mujica-Mota who provided feedback on the economic sections of the ERG report.

Rider on responsibility for report:

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

This report should be referenced as follows:

Greenhalgh J, Bagust A, Boland A, Blundell M, Oyee J, Dundar Y, Hockenhull J, Proudlove C, Chu P. Rituximab for the first-line maintenance treatment of follicular non-Hodgkin's lymphoma: A Single Technology Appraisal. LRiG, The University of Liverpool, 2010.

Contributions of authors:

Janette Greenhalgh	Project lead, drafted the clinical results section and supervised the production of the final report
Adrian Bagust	Critical appraisal of the economic model
Angela Boland	Critical appraisal of the economic evidence
Michaela Blundell	Critical appraisal of clinical statistical approach
James Oyee	Critical appraisal of clinical statistical approach
Yenal Dundar	Cross checking of manufacturer's search strategies
Juliet Hockenhull	Summary and critical appraisal of the clinical evidence
Chris Proudlove	Critical appraisal of the manufacturer's submission
Patrick Chu	Critical appraisal of the clinical sections of the manufacturer's submission

All authors read and commented on draft versions of the ERG report.

Table of contents

Abbreviations

AC	Appraisal Committee
AE(s)	adverse event(s)
AIC	Akaike Information Criterion
BIC	Bayesian Information Criterion
BSA	body surface area
CHOP	cyclophosphamide, doxorubicin, vincristine and prednisolone
CI	confidence interval
CR	complete response
CRu	complete response (unconfirmed)
CSR	clinical study report
CTX	chemotherapy
CVP	cyclophosphamide, vincristine and prednisolone
DSMC	Data and Safety Monitoring Committee
ECOG	Eastern Co-operative Oncology Group
EMA	European Medicines Agency
ERG	Evidence Review Group
ESMO	European Society for Medical Oncology
FACT-G	Functional Assessment of Cancer Therapy - General
FCM	cyclophosphamide, fludarabine, mitoxantrone
fNHL	follicular non-Hodgkin's lymphoma
HR	hazard ratio
HRQoL	health related quality of life
ICER	incremental cost-effectiveness ratio
IRC	Independent Review Committee
ITT	intention to treat
LYG	life years gained
MS	manufacturer's submission
MTE	median time to event
NICE	National Institute for Health and Clinical Excellence
OS	overall survival
PFS	progression- free survival
PD	progressive disease
PP	per protocol
PR	partial response
PRIMA trial	Primary Rituximab and Maintenance trial
PSA	probabilistic sensitivity analysis
QALY	quality adjusted life years
R-	rituximab with
RCT	randomised controlled trial
RTX	rituximab
SA	sensitivity analysis
SD	stable disease
SPC	Summary of Product Characteristics
STA	single technology appraisal
vs	versus
WTP	willingness to pay

1 SUMMARY

1.1 Scope of the submission

The remit of the Evidence Review Group (ERG) is to comment on the clinical and cost-effectiveness evidence submitted to the National Institute for Health and Clinical Excellence (NICE) as part of the single technology appraisal (STA) process. Clinical and economic evidence has been submitted to NICE from Roche in support of the use of rituximab (RTX) (MabThera®) as first-line maintenance treatment for patients with advanced follicular non-Hodgkin's lymphoma (fNHL). The manufacturer's submission (MS) describes the use of single agent RTX for patients whose disease has responded to induction therapy consisting of RTX + chemotherapy (R-CTX).

The current licence for RTX as a maintenance therapy is limited to use in patients with relapsed/refractory lymphoma. However, on 23rd September 2010, the European Medicines Agency's (EMA) Committee on Medicinal Products for Human Use issued a positive opinion to extend the use of RTX to include the treatment of follicular lymphoma patients responding to induction therapy. The final EMA decision is expected by 7th November 2010.

1.2 Summary of submitted clinical-effectiveness evidence

The main source of clinical evidence described in the MS is from the unpublished PRIMA trial. The PRIMA trial includes 1019 patients with high tumour burden, untreated fNHL (90% stage III/IV) who had a complete response (CR), unconfirmed complete response (CRu) or partial response (PR) to R-CTX induction therapy and were then randomised to receive either RTX monotherapy or observation for 2 years or until disease progression. The manufacturer reports clinical effectiveness results at several time points. At the data cut-off of June 2010, for the primary endpoint of progression-free survival (PFS), the median time to event (MTE) for patients in the RTX arm [redacted] At the data cut-off of January 2010, for the secondary endpoints of event-free survival (EFS), time to next anti-lymphoma treatment, time to next CTX treatment and overall response rate (RR), [redacted] For the outcomes of OS and transformation rate at first progression, [redacted] Subgroup analyses of age, gender, Follicular Lymphoma Prognostic Index (FLIPI) prognosis, type of induction CTX and response to induction [redacted]

[redacted]

1.3 ERG critique of clinical-effectiveness evidence

A single, international, multi-centre open-label, randomised controlled trial (RCT) (PRIMA) makes up the basis of the clinical and cost-effectiveness evidence in the MS. The results of PRIMA are as yet unpublished. Overall, the ERG considers that the PRIMA trial was well designed with centralised random allocation and safeguards to mitigate against possible bias. The treatments used in the PRIMA trial were similar to those used in UK clinical practice and the patients recruited to the trial were similar to those who would be treated in UK clinical practice. The ERG notes that several important protocol amendments (e.g. changing primary outcome and increasing sample size) were made during the course of the trial and the trial was stopped early; however, these actions appear to have been permitted in accordance with the protocol. The ERG notes that, at the time of closure, median PFS was not reached (and remains not estimable) and that none of the patients in the trial had been followed up for more than 4 years (recommended 7 years in the protocol). Whilst the clinical data appear to show a benefit for patients in the RTX arm, the ERG considers that data submitted are currently too immature to draw definite conclusions as to the clinical effectiveness of RTX as a first-line maintenance treatment for patients with fNHL.

1.4 Summary of submitted cost-effectiveness evidence

In the absence of any relevant UK-based economic evaluations of RTX as a first-line maintenance treatment for patients with fNHL, the manufacturer conducted a *de novo* economic evaluation. The manufacturer developed a four-state Markov model where patients are assumed to be within one of four possible discrete health states at any given time: "progression-free survival/first-line maintenance" (PFS1), "progression-free survival/second-line treatment" (PFS2), "progression" (PD) or "death". The model was developed to reflect a 25 year life-time horizon framework in order to capture all relevant costs and benefits associated with treating a patient with fNHL. The perspective adopted in the economic evaluation was that of the NHS and Personal Social Services (PSS) and costs and benefits were discounted at 3.5% per annum. Clinical-effectiveness data from two RCTs (PRIMA and EORTC 20981) were used to populate the submitted economic model. Quality adjusted life years (QALYs) were estimated using EQ-5D data from a previously published Roche-

commissioned study of UK patients with fNHL. In summary, the base case incremental cost effectiveness ratio (ICER) for first-line RTX maintenance therapy versus observation is estimated to be £15,978 per QALY gained. The manufacturer showed the ICER to be robust when subjected to structural, deterministic and probabilistic sensitivity analysis (PSA); the manufacturer's ICERs ranged from £8,966 per QALY gained to £21,151 per QALY gained.

1.5 ERG critique of cost-effectiveness evidence

The ERG is of the opinion that the cost effectiveness of RTX as a first-line maintenance therapy compared with observation in patients with fNHL who respond to first-line treatment cannot be assessed given the limited clinical data available.

The ERG identified a number of implementation issues, none of which appears to have a major impact on the estimated ICER for use of RTX as first-line maintenance therapy, either individually or in combination. Several of the more structural problems could not be corrected by the ERG, but it is unlikely that they will impact on any decision based on an assessment of cost effectiveness.

The ERG considers that lack of mature clinical effectiveness data from the PRIMA trial combined with the sensitivity of the base case ICER to several important model assumptions (e.g. age of the patients, assumed duration of benefit from RTX) indicate much greater uncertainty in model cost-effectiveness results than is suggested by the manufacturer's sensitivity analyses.

In summary, the ERG is of the opinion that the nature of the evidence submitted means that it is impossible to compare the cost effectiveness of first-line RTX maintenance vs observation in patients with fNHL with any confidence.

1.6 Commentary on the robustness of submitted evidence

1.6.1 Strengths

Clinical

The manufacturer provides clinical evidence from a single, unpublished, open-label RCT. The PRIMA trial is well-designed RCT. Additional information from the PRIMA trial was available to the ERG from the clinical study report (CSR) and the clarification responses provided by the manufacturer.

1.6.2 Weaknesses

Clinical and economics

The MS was lengthy (over 400 pages) and would have benefited from content-editing. There appears to be little cross-referencing between the clinical and economic sections of the MS in terms of clinical data utilised i.e. the economic evaluation is furnished with clinical effectiveness data from the EORTC 20981 trial which is not described in the clinical section of the MS.

The ERG notes that there is only one relevant RCT (PRIMA) describing RTX as a first-line maintenance therapy for this group of patients; this trial is of open-label design. The PRIMA trial was halted prematurely; the early stopping combined with a limited length of follow-up means that the data reported are immature. The ERG notes that it has been recently demonstrated that large differences in treatment effect sizes exist between trials that have been stopped early and similar trials that run their full course; this has been shown to be true regardless of the methodological quality of trials or the presence of statistical stopping rules.

The ERG considers that there have been too few events recorded in the PRIMA trial to make any firm conclusions regarding the effectiveness, and therefore cost-effectiveness, of RTX as a first-line maintenance treatment compared with observation alone for patients with previously untreated fNHL who have responded to R-CTX induction. The ERG notes that when the trial was stopped, none of the patients had been followed up for more than 4 years (the recommended follow up in the protocol was 7 years).

Economics

The ERG considers that there are a number of errors and inconsistencies in the economic model (e.g. omitted health states and pathways, use of PFS instead of EFS as key outcome, use of non-comparable clinical effectiveness evidence from a second clinical data source, non-standard use of discounting method, inaccurate costing of RTX treatment, underestimation of AEs). However, the ERG acknowledges that their correction would

probably not substantially affect the size of the manufacturer's base case ICER. The ERG is more concerned with the limited clinical trial data available to inform estimates of health gain in the economic model. Finally, the ERG has identified that the cost-effectiveness results are not robust to two key model assumptions (e.g. patient age and duration of RTX benefit).

1.6.3 Areas of uncertainty

Clinical and economics

The ERG considers that direct use of the submitted economic model results is too heavily affected by extensive and unquantifiable uncertainty around the central claims of clinical benefit to be useful for decision making i.e. the ERG concludes that the clinical data available make it impossible to compare the cost effectiveness of first-line RTX maintenance vs observation in patients with fNHL with any confidence.

The ERG is uncertain of the impact of RTX as a first-line maintenance treatment on the efficacy of second and third-line treatments. Follicular NHL is a treatable disease but it is not curable. Patients experience recurring and remitting disease over many years. Each successive relapse becomes more difficult to treat and it is not clear if there are optimal points in the treatment pathway for the use of RTX treatment in this patient group.

2 BACKGROUND

2.1 Critique of manufacturer's description of underlying health problems

In the context section of the MS (section 2) the manufacturer describes the key issues relating to the underlying health problem and associated risk factors. The MS provides an overview of the clinical problem, including aetiology, epidemiology and prognosis. A summary of this section is provided in Box 1 to Box 3, Table 1 and Table 2 . All information is taken directly from the MS unless otherwise stated.

Box 1 Aetiology

Follicular lymphoma is one of a group of diseases known collectively as non-Hodgkin lymphomas (NHLs) - cancers arising from the lymphoid cells of the immune system. These cells normally have a key role in protecting the body from pathogenic microorganisms. Malignant transformation of lymphocytes results in their uncontrolled replication. This usually starts within the lymph nodes, mainly those of the neck, armpits and groin. Swelling of these structures often provides the first clinical manifestation of illness, though other symptoms including fever, drenching night sweats, weight loss (so-called "B-symptoms") and tiredness may also be present at diagnosis or develop later.

Most cases of NHL, including fNHL, have no identifiable cause though a number of risk factors are known. These include chronic immunodeficiency caused by disease (e.g. rheumatoid arthritis and acquired immunodeficiency syndrome) or drugs (e.g. long-term antirejection therapy following organ transplantation), and certain infectious agents (e.g. Helicobacter pylori infection, which is associated with mucosa-associated lymphatic tissue lymphomas). Environmental factors such as occupational exposure to tetrachloroethylene and agricultural biocides have also been suggested as possible causes of lymphoma[1]

Box 2 Epidemiology

In 2007, there were 10,917 new cases of NHL recorded in England and Wales;[1] 22% (2,024) of these were likely to be fNHL.[2] Of these, 85% (1,720) would have been diagnosed as stage III/IV cancers requiring systemic therapy.[3]
The age-standardised incidence rate for NHL increased by more than a third (35%) in the twenty-year period between 1988-2007, mirroring the increases in many other countries.[1] This increase in incidence is considered genuine rather than a result of improved diagnosis and that it is the consequence of an increase in the incidence of several lymphoma subtypes, including fNHL.[1] In the late 1990's and early 2000's incidence showed signs of levelling off suggesting that the recent rapid growth in new cases of NHL is slowing.
The incidence of NHL is similar in men and women (M/F ratio 1.1/1.0) and increases with age – rates increase sharply in people over 50 and around two-thirds of all cases are diagnosed in people over 60 years of age.[1]

Box 3 Prognosis

Survival for patients with fNHL is prolonged. Different figures for median survival have been reported, but 8-10 years from diagnosis is typical.[4, 5] However, these are likely to be underestimates since there is good evidence from recent large population-based[6] and single institution studies that survival is improving, probably as a consequence of improved treatment.[7-9] Even at 8-10 years, survival is about double that reported in the years before the advent of cytotoxic CTX indicating that appropriate treatment does alter the long-term course of the disease.[10] Despite this, most patients with fNHL ultimately die of their disease. For example, amongst a group of 147 patients followed for over 15 years from diagnosis by Lister, 94 died during the observation period, with 76 deaths attributed to progressive lymphoma.[5]

Prognosis is partly determined by the extent of disease at diagnosis, which is usually described using the Ann-Arbor[11] staging system.

Other factors besides disease stage have been identified as having prognostic significance. Five of these were incorporated into the International Prognostic Index (IPI) which allows a composite IPI score to be calculated.[12] Although the IPI was formulated for aggressive lymphomas it was also applied to more indolent forms of the disease, like fNHL. More recently, the Follicular Lymphoma Prognostic Index (FLIPI)[13] has been devised specifically for this type of lymphoma. Although the FLIPI is well accepted as having prognostic significance, it is not routinely used to guide treatment, which is generally determined by disease stage plus clinician and patient preference for a particular CTX regimen.

Table 1 Ann-Arbor staging system of NHL

Stage 1	Involvement of a single lymph node region (I), or localised involvement of a single extralymphatic organ or site (IE)
Stage II	Involvement of two or more lymph node regions on the same side of the diaphragm (II), or localised involvement of a single associated extralymphatic organ or site and its regional nodes with or without other lymph node regions on the same side of the diaphragm (IIE)
Stage III	Involvement of lymph node regions on both sides of the diaphragm (III), that may also be accompanied by localised involvement of an extralymphatic organ or site (IIIE), by involvement of the spleen (IIIS), or both (IIIE+S)
Stage IV	Disseminated (multifocal) involvement of one or more extralymphatic organs with or without associated lymph node involvement, or isolated extralymphatic organ involvement with distant (nonregional) nodal involvement.

The FLIPI[13] index (used for prognosis purposes) is noted by the manufacturer (see above) but is not described in the MS; Table 2 is provided by the ERG for reference. The FLIPI[13] index consists of five significant risk factors prognostic of overall survival (OS). These include: age (<60 years or >60 years); Ann Arbor Stage III or IV; haemoglobin level <120g/L; elevated serum lactate dehydrogenase; >4 nodal sites. The number of prognostic factors impacts on OS, so that the lower the number of risk factors, the greater the chances of survival at 5 and 10 years.

Table 2 FLIPI risk scores

Risk group	Number of factors	% 5 year OS	% 10 year OS
Low	0-1	90.6	70.7
Intermediate	2	77.6	50.9
High	3+	52.5	35.5

The MS provides an adequate description of the underlying health problems in patients with fNHL. For clarity, the ERG would like to add that fNHL is a CD20+ cancer of the B lymphocytes. These lymphomas are low grade and not curable, but tend to develop slowly (although they can 'transform' into high grade lymphomas that may be more difficult to treat). There exists a histological grading system for fNHL that enables classification into Grades I, II or III, with III subdivided into a and b. Grade IIIb follicular lymphomas are generally treated as for high grade lymphoma.

With reference to the figures stated in the epidemiology section of the MS (MS, p.32) the ERG notes from the final scope issued by NICE[14] that the number of follicular lymphomas as a percentage of all NHLs will differ according to the classification system used, between 22% and 40%; a figure of 22% is used in the MS. It is worth noting, that in Britain the biggest increases in incidence of NHL have been in older people and the rates in those over 75 are three times higher than they were in 1975.[1]

2.2 Critique of manufacturer's overview of current service provision

The MS provides a reasonable description of current service provision, including first-line treatment options (Box 4). First-line CTX options are described in Box 5. There is an introduction to the concept of maintenance treatment in patients with advanced fNHL who have responded to first-line R-CTX (Box 6) and discussion of the possible impact of RTX first-line maintenance on the efficacy of subsequent treatments (Box 7).

Box 4 First-line treatment options for fNHL

Stage I/II
Approximately 15% of patients present with early-stage disease.[3] Patients considered to have limited, stage I-II, disease may be candidates for localized radiotherapy treatment, which can have a curative potential.[15] Around half of patients so treated are free of relapse after 5 years. Patients who reach this point have a very low risk of future relapse. In one large series, relapse-free survival rates were 55%, 44%, 40% and 37% at 5, 10, 15 and 20 years, respectively, suggesting that only a subpopulation of patients will have a prolonged disease-free interval after radiotherapy, but that for this group, relapse more than 10 years after treatment is rare.[16]

Stage III/IV
Systemic therapy is only recommended in patients with stage III and IV disease with evidence of systemic symptoms, high tumour burden, rapid disease progression, or other key features. To date, no curative therapy had been established for patients with advanced follicular lymphoma,[15] so the natural course of the disease once treatment is required follows a cycle of treatment induced remission followed by eventual relapse – with each remission duration becoming shorter and less patients responding to each cycle of therapy.[17, 18]

Antibody monotherapy or single agent alkylating agents (e.g. chlorambucil) can be considered an alternative in previously untreated fNHL patients with particularly low risk disease, or those unsuitable for more intensive treatments.[15]

Box 5 First-line CTX options

Approximately 93% of all eligible previously untreated stage III-IV follicular lymphoma patients in the UK currently receive R-CTX as standard treatment.[19] The lack of consensus in terms of the preferred combination partner for RTX in this setting is likely driven by several factors, including (i) a breadth of data from several randomised trials and a meta-analysis[20-26] demonstrating that the clinical benefit associated with the addition of RTX to CTX is independent of the CTX backbone. European guidelines currently specify that if complete remission and long progression free survival is to be achieved, rituximab plus CTX (R-CTX), usually cyclophosphamide, doxorubicin, vincristine and prednisolone (CHOP), cyclophosphamide, vincristine and prednisolone (CVP) or cyclophosphamide, fludarabine, mitoxantrone (FCM) or bendamustine should be used in advanced stage fNHL patients requiring treatment.[15] Current NICE guidelines on the use of RTX in previously untreated fNHL (TA110) only specify the use of R-CVP in this setting, as was consistent with the marketing authorization for RTX at the time of review in 2006.[27] The licence for RTX has since been broadened to accommodate its use in combination with any CTX, based on more recent evidence and is reflected in the European guidelines[15] and Scottish Medical Consortium guidance.[28]

Box 6 Concept of maintenance treatment

Maintenance therapy is not currently widely used after patient's first response to immunochemotherapy, and standard practice in the UK is to closely observe patients until evidence of disease progression, at which point second-line therapy is initiated.
The aim of using RTX maintenance therapy in previously untreated fNHL patients responding to induction therapy is to extend and deepen the first and often most durable remission.[17, 18] It would be expected, as in the relapsed/refractory follicular lymphoma setting, that extended progression free survival yielded by RTX maintenance would delay time to relapse and therefore the burden of further CTX treatment.[29, 30] As demonstrated in the relapsed setting this approach may also ultimately extend patient OS but will require a significantly longer period of follow-up to become apparent.

The MS provides an outline of the current treatment pathway for fNHL. This is reproduced in Figure 1. The proposed place for RTX maintenance treatment is as a replacement for observation following successful induction with R-CTX. The ERG notes that the second-line option recommended by NICE[31] is induction with R-CTX or RTX monotherapy followed by RTX as a maintenance treatment (for patients who respond to induction).

The manufacturer estimates that 1656 patients per year will be eligible for treatment with first line RTX maintenance (MS, p.31). This figure is uncertain as it is based on the manufacturer's assumptions that i) 22% of all cases of NHL are diagnosed as fNHL and ii) there is a response rate of 90% to first-line induction with R-CTX (derived from results of pivotal PRIMA[32] trial reported in the MS).

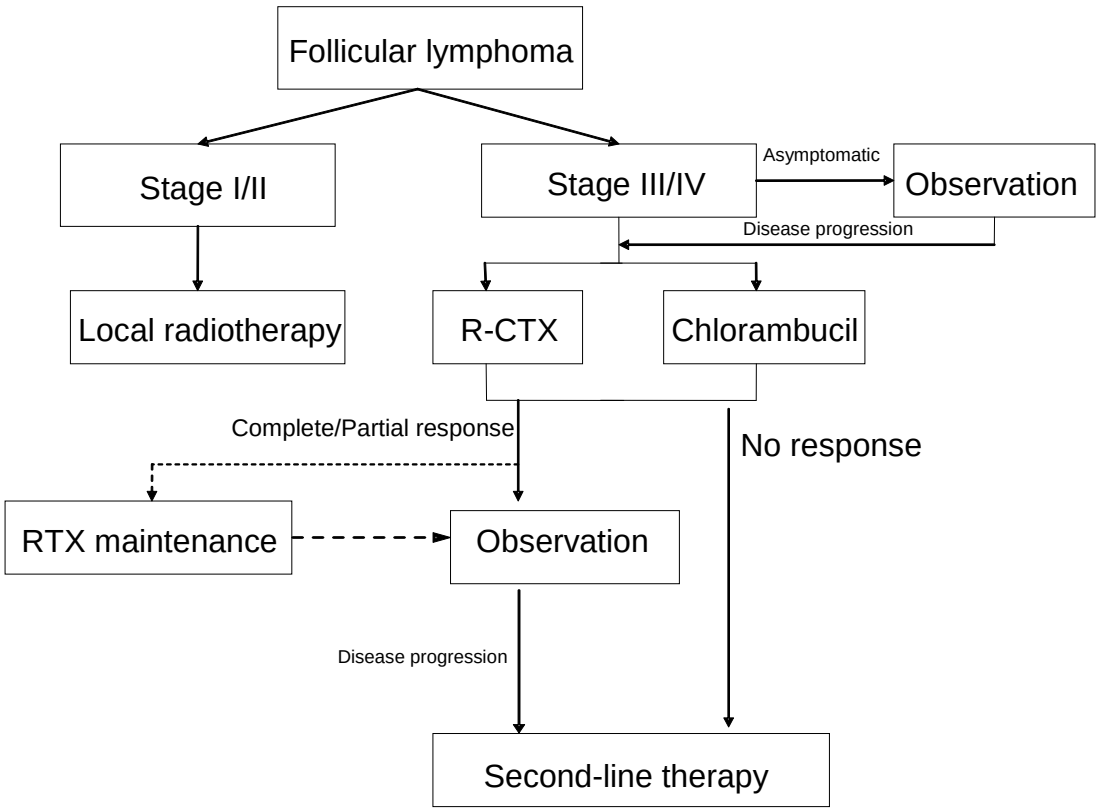

Figure 1 Outline of fNHL treatment pathway in England and Wales

Box 7 Impact of first-line maintenance on subsequent treatments

.....it is likely that patients who have had a "durable" period of remission after finishing maintenance therapy will qualify for re-treatment with R-CTX induction. Patients subsequently responding to second-line R-CTX are also likely to qualify for re-treatment with RTX maintenance. Current European Society for Medical Oncology (ESMO) guidelines[15] state that in terms of the treatment of relapsed disease"…RTX should be added if the previous antibody containing scheme achieved a >6-month duration of remission" and "RTX maintenance for up to 2 years has a favourable side-effect profile and based on a systematic meta-analysis, substantially prolongs progression-free survival (PFS) and overall survival (OS) in relapsed disease even after antibody-containing induction.[15]

Several clinical trials also support the concept of re-treatment with RTX at relapse,[33-36] suggesting that re-treatment with either RTX monotherapy or in combination with CTX can induce equally long or even longer response durations that those obtained following the initial use of RTX. Finally, NICE have also considered the concept of RTX re-treatment in TA137.[31] The Appraisal Committee in its consideration of the evidence heard from clinical experts that "the evidence indicated that follicular non-Hodgkin's lymphoma could be re-treated with RTX with little or no loss of efficacy. Although it noted this as an area of uncertainty, the Appraisal Committee accepted that this was biologically plausible given its (RTX's) mechanism of action"[31]

The ERG is satisfied with the manufacturer's description of current service provision and reinforces the manufacturer's characterisation of stages III/IV fNHL; i.e. that it is incurable with current treatments focused on reducing symptoms and minimising toxicities. Patients can experience periods of prolonged remission, but will inevitably relapse and require subsequent treatments. With each successive treatment the chances of remission are lower and the duration of the remissions achieved, shorter.[37]

3 CRITIQUE OF THE DECISION PROBLEM

In the MS, the manufacturer presents the decision problem issued by NICE,[14] and the manufacturer's rationale for any deviation from this in the MS. This is reproduced in Table 3.

Table 3 Decision problem as addressed in MS

	Final scope issued by NICE	Decision problem addressed in the MS	Rationale if different from scope
Population(s)	Adults with advanced fNHL that has responded to first-line CTX	Adults with advanced follicular lymphoma that has responded to first-line R-CTX	R-CTX induction therapy is the current gold standard for previously untreated fNHL patients in the UK, with approx 93% of eligible patients receiving this treatment option. The vast majority of patients not treated with R-CTX, receive chlorambucil monotherapy (approx 5% of all eligible first-line fNHL patients). These patients tend to be older, frailer, and with co-morbidities that make them ineligible for treatment with either R-CTX or RTX maintenance therapy.
Intervention(s)	RTX maintenance therapy	As final scope	
Comparators	• Standard management without RTX maintenance therapy • Ibritumomab tiuxeten (IT) (zevalin®)	Standard management without RTX maintenance therapy (i.e. observation)	• No evidence to support clinical benefit of zevalin in previously untreated advanced fNHL patients induced with R-CTX • Minimal zevalin usage in UK.
Outcomes	• Progression-free survival • Overall survival • Response rates • Adverse effects of treatment • Health-related QoL	As final scope	
Economic analysis	• The reference case stipulates that the time horizon for estimating clinical and cost effectiveness should be sufficiently long to reflect any differences in costs or outcomes between the technologies being compared		Due to extensive censoring of overall survival in PRIMA (95% and 84% in the rituximab and observation arms, respectively), the probability of progressing or the probability of dying in second line or third line were obtained from the EORTC 20981trial (6 years median follow-up)
Other considerations	Guidance will only be issued in accordance with the marketing authorisation. If evidence allows, the following subgroups will be considered: • whether RTX was received in combination with first-line CTX • type of first-line chemo-immunotherapy regimen received • type of response (that is, CR vs PR) achieved after first-line treatment	See rationale	Given the revised patient population in the decision problem (i.e. patients responding to first-line R-CTX), the "subgroup" addressed in the first bullet is no longer relevant. Please note, the PRIMA trial, which forms the core of this MS, did not include a sub-population of patients induced with CTX alone. Instead, all patients received standard first-line therapy for previously untreated fNHL: R-CTX. The potential impact of baseline demographics and prognostic factors on the treatment effect in the PRIMA study was assessed by analysing the following subgroups (non-randomised) and will be addressed in the MS: age (≥ 60 years, < 60 years), gender (male, female), pre-induction FLIPI score (≤ 1, 2, ≥ 3), induction treatment (R-CHOP, R-CVP, R-FCM), and response to induction treatment (CR/CRu, PR). Given the revised patient population in the decision problem (i.e. patients responding to first-line R-CTX), the "subgroup" addressed in the first bullet is no longer relevant.
Related NICE Guidance Technology Appraisal No.137, February 2008 (review of technology appraisal No. 37), "Rituximab for the treatment of relapsed or refractory stage III or IV follicular non-Hodgkin's lymphoma". Review date December 2010. Technology Appraisal No. 110, September 2006; "Rituximab for the treatment of follicular lymphoma". Review date June 2009. Clinical Guideline No. CSGHO, October 2003, "Improving outcomes in haemato-oncology cancer" (expected review date TBC).			

3.1 Population

In the decision problem issued by NICE, the population in the final scope[14] is defined as 'adults with advanced fNHL that has responded to CTX'.

The patients with fNHL in the PRIMA[32] trial are described as having a 'high tumour burden'. The patient characteristics listed in the MS include the patient's disease stage and shows that 90% of patients had disease at stages III or IV whilst 10% of patients were classified as having disease at stages I or II. Since disease at stages III and IV is classified as advanced and disease at stages I and II are generally treated with radiotherapy rather than CTX, the ERG requested clarification from the manufacturer as to why patients with disease at stages I and II were included in the PRIMA[32] trial. The manufacturer's response included the following statement: 'In the small proportion of patients with limited stage I-II disease, radiotherapy (involved or extended field, 30–40 Gy) is the preferred treatment option having a curative potential. However, in stage I-II patients with large tumour burden systemic therapy may be applied as indicated for advanced stages (i.e. with the occurrence of symptoms including B symptoms, haematopoietic impairment, bulky disease, vital organ compression, ascites, pleural effusion or rapid lymphoma progression)'. The manufacturer also provided statements to this effect from three clinicians working in UK clinical practice. These confirm that for some patients with disease at stages I or II, treatment with CTX is standard UK practice.

The ERG notes that all the patients in the PRIMA[32] trial received R-CTX induction therapy; none were treated with CTX alone. The manufacturer has chosen to restrict their consideration of the decision problem to those patients who had received first-line treatment with R-CTX as it considers that R-CTX is the standard treatment used in clinical practice in the UK; the MS states that the 5% of total population not offered R-CTX will be given chlorambucil monotherapy: 'These patients tend to be older, frailer, and with co-morbidities that make them ineligible for treatment with either R-CTX or RTX maintenance therapy' (MS p.42). The ERG's consultations with clinical experts confirm that this is the case.

3.2 Intervention

Rituximab does not currently have a UK marketing authorisation for the indication detailed in the MS (i.e. as maintenance therapy in previously untreated fNHL patients responding to induction with R-CTX). However, on 23rd September 2010, the European Medicines Agency's (EMA) Committee on Medicinal Products for Human Use issued a positive opinion to extend the use of RTX to include the treatment of patients with FL who respond to induction therapy. The final EMA decision is expected to be known by 7th November 2010.

3.3 Comparators

The decision problem issued by NICE in its final scope[14] states that RTX maintenance treatment should be compared with standard management without CTX treatment (observation) or zevalin.

In the MS, the only comparator considered is observation. The manufacturer states that in the UK, the current standard management of patients with fNHL is observation following first-line induction therapy if patients respond to first-line induction. The ERG confirms that this is the case.

The manufacturer has dismissed zevalin as a comparator claiming: i) that there is no clinical evidence to support the clinical benefit of zevalin in patients with untreated advanced fNHL and ii) there is minimal zevalin use in UK clinical practice. With regard to i) the ERG notes that the FIT[13] trial supports the use of zevalin as a consolidation therapy for patients with advanced fNHL who had responded to induction CTX; however of the 414 patients included in the trial, the majority (86%) received induction CTX that did not include RTX. The published paper[13] states that the subgroup analysis of the 14% of patients who did receive induction R-CTX demonstrated a statistically significant difference between the intervention and control groups on the outcome of PR-CR conversion rate, but not in PFS. It is noted by the trial authors that the trial was not designed or powered to detect differences in PFS outcomes based on first-line induction treatments and that the patients treated with R-CTX were included in the latter stages of the trial and so had shorter follow-up. The ERG is also aware that the Summary of Product Characteristics (SPC) for zevalin states that zevalin is indicated as consolidation therapy after remission induction in previously untreated patients with fNHL, however the benefit of zevalin following R-CTX has not been established.[38]

With regard to the manufacturer's assertion that there is minimal use of zevalin in clinical practice in the UK, the MS provides IMS hospital usage data for zevalin in the UK. This is reproduced in Table 4. According to the MS, the reported expenditure of £34,836 in 2009 'equates to a total of 5 patients being treated with zevalin (across all indications) over this period, assuming that one dose of zevalin in administered per patient.'[39]

Table 4 IMS hospital usage for zevalin in UK

	2006	2007	2008	2009	Year to date: May 2010
zevalin	£7,250	£29,000	£43,500	£34,836	£0

The ERG considers that the manufacturer makes a convincing case for not including zevalin as a comparator.

3.4 Outcomes

The manufacturer has addressed all the outcomes stated in the scope issued by NICE,[32]including progression-free survival (PFS), overall survival (OS), response rate (RR), adverse events (AEs) and quality of life (HRQoL); however, the ERG considers that the data relating to these outcomes are not yet mature enough to inform decision-making and therefore no reliable conclusions can be drawn from the results of the PRIMA[32] trial regarding the efficacy of RTX as first-line maintenance treatment compared with observation in patients with fNHL who have responded to induction therapy.

3.5 Economic analysis

In the manufacturer's response to the decision problem defined by NICE in its final scope[14] the manufacturer states that its economic case is partly based on the outcomes of the EORTC 20981[29, 30] trial (Table 3)

> *Due to extensive censoring of overall survival in PRIMA (95% and 84% in the rituximab and observation arms, respectively), the probability of progressing or the probability of dying in second line or third line were obtained from the EORTC 20981 trial (6 years median follow-up).*

The manufacturer has not described the characteristics of the EORTC 20981[29, 30] trial in the clinical section of the MS, neither have the merits or otherwise of using the trial been discussed by the manufacturer. The ERG considers that this is a major omission in the MS.

The results of the EORTC 20981[29, 30] trial were originally published in 2006, but were updated in a second publication in 2010[30] that included data from a 6 year follow-up period. The trial included patients with relapsed/resistant fNHL (n=465) randomised to treatment with either R-CHOP or CHOP alone. Those who achieved a CR or PR (n=334) were randomised again to either maintenance treatment with RTX (once every 3 months) or observation for 2 years or until relapse. Maintenance treatment with RTX significantly improved PFS compared with observation (median 3.7 years vs 1.3 years). Five-year OS was not significantly different between the arms, (74% in RTX arm and 64% in the observation arm). The authors suggest that this lack of difference may in part, be due to the 'unbalanced use of RTX in post-protocol salvage treatment'. Maintenance with RTX was associated with statistically significant increases in grade 3 and 4 infections.

The characteristics of the EORTC 20981[29, 30] trial are described in Table 33 (Appendix 5). The trial is considered to be of good quality (Table 34, Appendix 6). The manufacturer has used the patient outcome data from the EORTC 20981[29, 30] trial to predict the long-term outcomes of patients in the PRIMA[32] trial. The ERG considers that this extrapolation may not

be valid in that the patients in the EORTC 20981[29, 30] trial are dissimilar to those in PRIMA[32] in terms of prior treatment histories.

Patients were included in the EORTC 20981[29, 30] trial if they had relapsed following two previous non-anthracycline-containing CTX regimens. In addition, patients were RTX naive. This means that none of the patients were treated with CHOP or R-CHOP as a first-line CTX. Although accepted practice at the time when the EORTC 20981[29, 30] trial was conducted, the lack of CHOP or R-CHOP at first-line would now be considered as sub-optimal treatment. In contrast, the majority of patients in the PRIMA[32] trial (74%) received R-CHOP as their first-line treatment (all patients received R-CTX). This is in line with current clinical practice in the UK. Further, the patient baseline characteristics document that 50% of patients in the EORTC 20981[29, 30] trial were randomised into the trial more than 2 years after their initial diagnosis. This was not the experience of patients in PRIMA.[32] For these reasons, the ERG is not confident that the outcomes of the patients in the EORTC 20981[29, 30] trial are a true representation of the future outcomes of the patients in the PRIMA[32] trial.

3.6 Other considerations

In the decision problem, NICE has suggested that if the evidence allows, three subgroups of patients should be considered, including i) whether RTX was received in combination with first-line therapy; ii) the type of first-line chemo-immunotherapy received and iii) type of response (CR/PR) after first-line treatment. As noted above, in the pivotal PRIMA[32] trial all patients were treated with R-CTX induction therapy; the manufacturer claims that i) is therefore redundant. The manufacturer has considered ii) and iii) in its subgroup analyses of the PRIMA[32] trial; however, given that the trial was not powered to show significant differences between subgroups and given the ERG's concerns regarding the prematurity of the dataset, the results of any subgroup analyses should be treated with caution.

3.7 Related NICE guidance

Current NICE guidance[31] recommends R-CTX as a second-line induction therapy for patients with relapsed/refractory fNHL. Patients who respond to induction may then be treated with RTX as a maintenance treatment. However, patients in the pivotal maintenance trial (EORTC 20981[29, 30]) were RTX-naive i.e. had not received RTX as a first-line induction and/or maintenance therapy. This means that if RTX is approved by NICE for first-line maintenance, then the impact of this decision on current NICE guidance on second-line treatments for this group of patients needs to be re-considered.

As reported in the MS (p.36) and included in the background section of the ERG report, the manufacturer discusses the issue of re-treatment with RTX. The manufacturer states that patients who have a durable period of remission after finishing maintenance therapy will qualify for re-treatment with R-CTX induction; patients responding to second-line induction are eligible for re-treatment with RTX maintenance therapy. In support, the manufacturer cites European Society for Medical Oncology (ESMO) guidelines,[15] the Appraisal Committee's decision regarding TA137[31] and four studies[33-36] that examined the effects of re-treatment with RTX.

The ERG notes that the ESMO guidelines[15] state that salvage treatment depends on efficacy of prior regimens. In early relapses (<12 months), a non-cross resistant scheme should be preferred (e.g. fludarabine after CHOP). Rituximab should be added if the previous antibody-containing scheme achieved a > 6-month duration of remission. The guidelines[15] also state that RTX maintenance for up to 2 years has a favourable side-effect profile and based on a systematic meta-analysis, substantially prolongs PFS and OS in relapsed disease even after antibody-containing induction.

In the consideration section of the guidelines for TA137,[31] it is recorded that the AC accepted that re-treatment with RTX 'with little or no loss of efficacy' was biologically plausible given RTX's mechanism of action. The MS (p.36) also describes a number of studies that the manufacturer maintains 'support the concept of re-treatment with RTX at relapse.' The ERG notes that the studies outlined in Table 5 all conclude that re-treatment with RTX is effective and safe; however none of the patients in these studies had received RTX first-line maintenance treatment as proposed in the MS, followed by RTX as second-line maintenance treatment as recommended in TA137.[31]

Table 5 Clinical trials using RTX as re-treatment

Study	No evaluable	Disease	Re-treatment regimen
Johnston[35]	178	B-NHL	RTX x4 R-CTX
Igarashi [34]	13	IL	RTX x4
Davis[33]	58 (57)	FL=95% SLL=5%	RTX x4
Coiffier[36]	59	B-NHL	RTX x4 R-CTX

TTP = time to progression; RD = response duration; IL = indolent lymphoma; FL: follicular lymphoma; SLL = small lymphocytic lymphoma

4 CLINICAL EFFECTIVENESS

4.1 Critique of manufacturer's approach

Table 6 provides an outline of the key background/clinical information and its location within the MS. Its purpose is to signpost the reader to the main areas of background/clinical information within the MS.

Table 6 Key clinical information in the MS

Key information	Pages in the MS
Description of the technology	20
Statement of decision problem	42-45
Context	28-40
Equity and equality	41
Literature search	47-49
Search strategies	326
Study selection	49-55
Clinical effectiveness evidence:	
Trial information	58-96
Results: main and subgroups	96-136
Results: updated main and subgroups	145-148
Results: quality of life	137-145
Results: safety	153-195
Results: updated safety	209-211
Results: laboratory parameters	196-208

4.1.1 Description of manufacturers search strategy and comment on whether the search strategy was appropriate.

The manufacturer describes the literature searches carried out by a third party in February 2010; these searches were subsequently updated by the manufacturer so as to include any further publications up until July 2010. The ERG is confident that all major electronic databases were searched including the Cochrane Library, Ovid Medline (1950 to present), Medline In Process, Ovid Embase (1980 to present) and Database of Clinical trials in Haematological Malignancies. Appropriate hand searching was conducted to identify any additional studies, this included conference proceedings (from 2005 to 2010) of: European Haematology Association (EHA), American Society of Haematology (ASH), American Society of Clinical Oncology (ASCO) and International Society for Pharmacoeconomics and

Outcomes Research (ISPOR). The reference lists of previous trials and systematic reviews were also searched.

The MS provides a clear description of the searches carried out to identify primary relevant research. The comprehensive search strategy used drug names and disease area, a filter for controlled trials and adopted no language restrictions. In line with the manufacturer's stated aim, the search included other treatments that might be used as maintenance or consolidation therapy in this setting.

4.1.2 Inclusion/exclusion criteria used in the study selection

The MS provides a detailed report of the inclusion/exclusion criteria applied to the selection of potentially relevant studies. These are described in Table 7.

Table 7 Inclusion/exclusion criteria

	Inclusion	**Exclusion**
Population	Patients with stage III/IV follicular non-Hodgkin lymphoma responding (CR or PR) to first-line R-CTX induction treatment, thus eligible for maintenance.	Patients that do not demonstrate stage III/IV follicular non-Hodgkin lymphoma Patients not receiving RTX maintenance after responding (CR or PR) to first-line R-CTX induction treatment
Intervention	RTX maintenance following response (CR or PR) to first-line R-CTX	Studies not including RTX maintenance vs observation after response (CR or PR) to first-line R-CTX induction therapy Studies without R-CTX as first-line induction therapy
Comparators	Observation following response (CR or PR) Other agents used as maintenance or consolidation following response (CR or PR) to first-line R-CTX. These include: Oral alkylating agents: chlorambucil/cyclophosphamide Chemotherapeutic agents: fludarabine/bendustamine/oblimersen Other monoclonal antibodies: Alemtuzumab/ofatumumab Radioactive monoclonal antibodies: ^{90}Y-ibritumomab tiuxetan/tositumomab Vaccines: BioVazID/Oncoquest-L Other agents: Interferon 2-alpha Any combination regimen containing at least one of the above	
Outcomes	Clinical outcomes reported at the end of maintenance therapy Response to treatment (overall response, CR, PR) OS EFS Time to first progression Time to re-treatment (time to next CTX or time to next anti-lymphoma treatment) Transformation rate Therapy-related morbidity and mortality FLIPI index AEs Withdrawals	None
Study design	Phase III RCTs Meta-analyses and systematic reviews of Phase III RCTs No language restrictions	Studies that were not randomised, controlled, phase III clinical trials, or reviews or meta-analyses of such trials

The ERG is satisfied with the review process as described in Appendix 1 of the MS (MS, p.341).

4.1.3 Table of identified studies

The search conducted by the manufacturer identified one study for inclusion in the review, the PRIMA[32] trial. The results of the interim analyses of the PRIMA[32] trial were presented at two international conferences (ASCO and EHA) in 2010; however, the ERG notes that full publication of the trial results is not expected until 2011. For the purposes of this ERG report, the Clinical Study Report (CSR)[32] is used to reference the PRIMA[32] trial.

An appropriate QUORUM[40] flow diagram describing the review process is provided by the manufacturer (MS, p53). It indicates that 30+ publications were identified that reported outcomes for trials employing maintenance therapy with a variety of treatments (RTX, CTX, interferon, zevalin). However, none of these publications were included in the final review stage. Since the MS does not provide a list of excluded publications, the ERG is unable to verify whether these were rightfully excluded.

4.1.4 Details of relevant studies not included in the submission

The ERG is confident that the PRIMA[32] trial is most relevant to the manufacturer's specified criteria; however, other trials might be considered relevant to the decision problem issued by NICE.[14] The ERG notes that the manufacturer's omission of CTX therapy alone as induction treatment will have impacted on the included trials and any RCTs with CTX induction therapy followed by RTX maintenance treatment will have been excluded. For example, the ECOG 1496[38] trial appears to be relevant to the decision problem. Patients (n=282) with previously untreated fNHL were randomised to CVP induction CTX. Responders (n=228) to induction therapy were randomised to RTX maintenance or observation. The study was terminated early when the primary endpoint of PFS was crossed in favour of the maintenance arm. Median PFS was 4.2 years in RTX maintenance arm compared to 1.5 years in the observation arm.

4.2 Description of the included study

The PRIMA[32] trial is:

> *an open-label, international, multicenter, randomised trial with two treatment phases. During the first phase ('induction phase'), patients with advanced follicular lymphoma were evaluated for response to one of three different rituximab plus CTX induction regimens (R-CVP, R-CHOP, R-FCM). Patients who responded to induction treatment (ie, achieved a confirmed or unconfirmed complete response [CR/CRu] or partial response [PR]) were randomised to receive either rituximab maintenance therapy (one dose every eight weeks for two years, for a total of 12 doses) or no further treatment (observation) in the second phase of the study ('maintenance/observation phase'). All randomised patients were treated or observed for two years or until disease progression, whichever occurred first. All patients who completed the maintenance/observation phase were then followed up for a further 5 years.(MS, p. 59)*

The trial characteristics of PRIMA[32] are described in Table 8.

The ERG emphasises that the full clinical paper describing the PRIMA[32] trial is not expected to be submitted for publication until 2011. Full details of the inclusion/exclusion criteria used in PRIMA[32] are presented in the MS (p.61-63).

The key outcome measures and terms used within the PRIMA[32] trial are described in Table 30 of Appendix 2. Assessments related to progression were based on the criteria for evaluation of response in NHL.[41] Response to induction therapy was assessed after 3 (R-FCM) or 4 (R-CVP/R-CHOP) cycles of combined immuno-chemotherapy. Patients with a CR, CRu, or PR, continued their designated course of treatment. Patients with disease progression were withdrawn from the trial as were patients with stable disease. Final response assessments were carried out at the end of the induction phase; patients with CR, CRu or PR were randomised into the maintenance phase. During maintenance, assessments related to progression were undertaken at each clinic visit, along with ECOG performance status scores.

Quality of life (QoL) was assessed using the EORTC QLQ-C30[42] (version 3) and the Functional Assessment of Cancer Therapy–General[43] (FACT-G, version 4.0). These questionnaires were distributed by the investigators to patients at screening, at the end of induction therapy, and then every year during the maintenance/observation and follow-up period (i.e. potentially a total of nine questionnaires per patient) until documented disease progression. Patients were encouraged to complete the questionnaire within one week of the respective study visit.

Table 8 PRIMA trial characteristics

Study	Trial design and patients	Intervention/ comparator	Key inclusion criteria	Key exclusion criteria	Outcomes
PRIMA[32]	• RCT, Phase III, international, open-label • 24 countries including UK (16 patients in 9 centres) • Previously untreated patients with high tumour burden fNHL • Patients achieving CR/CRu or PR following induction R-CTX were randomised to receive either RTX monotherapy as maintenance, or observation, for 2 years or until disease progression All RTX doses 375mg/m^2	<u>Induction treatment:</u> N= 1202 R-CHOP (6 cycles) n= 885 R-CVP (8 cycles) n=272 R-FCM (6 cycles) n=45 All RTX = 8 treatments <u>Maintenance RTX</u> N=506 Infusion every 8 weeks for 2 years or until progression <u>Observation</u> <u>N=513</u> Observation every 8 weeks for 2 years or until progression	<u>Induction phase</u> • Histologically confirmed fNHL grades 1,2 or 3a, with lymph node biopsy performed within 4 months before study entry and with material available for central review • Previously untreated fNHL • At least one high tumour burden GELF criterion requiring initiation of treatment • >18 years • ECOG ≥2 • Adequate haematological function • Women: not breast feeding, using effective contraception, not pregnant and agreed not to become pregnant during trial or 12 months thereafter • Men: agreed not to father a child during trial or 12 months thereafter <u>Maintenance phase</u> • CR/CRu/PR to R-CTX induction • All indicator lesions reported on the on-study form must have been re-evaluated	<u>Induction phase</u> • Transformation to high-grade lymphoma • Grade 3b fNHL • Presence/history central nervous system disease • Regularly taking corticosteroids during 4 weeks prior to study entry (≥ 20 mg/day prednisone) • Prior or concomitant malignancies • Major surgery within 28 days prior to registration • Poor renal or hepatic function • HIV infection, active hepatitis B or C • Serious underlying medical conditions that could impair ability to participate • Life expectancy >6 months • Sensitivity/allergy to murine products • Other co-existing medical or psychological condition that would preclude participation <u>Maintenance phase</u> • Serious underlying medical conditions that could impair ability to participate • Did not complete all cycles of induction treatment due to toxicity/had not completed at least four cycles of R-CHOP + 2RTX, six cycles of R-CVP, or four cycles of R-FCM induction treatment • Delay in treatment ≥14 days following any cycle of induction CTX	<u>Primary</u> PFS <u>Secondary</u> • EFS • OS • Time to next anti-lymphoma treatment • Time to next CTX treatment • Overall RR at the end of maintenance/observation • Transformation rate at first progression • QoL • Safety

ECOG= Eastern Co-operative Oncology Group; GELF= Groupe d'Etude des Lymphomes Folliculaires;

4.3 Description and critique of manufacturers approach to validity assessment

A single RCT (PRIMA[32]) makes up the basis of the clinical and cost-effectiveness evidence in the MS. The results of the PRIMA[32] trial are, as yet, unpublished. This section outlines the strengths and weaknesses of the PRIMA[32] trial. Data in this section are taken from the MS as well as from data subsequently provided by the manufacturer as a part of the STA clarification process.

In response to a clarification request from NICE, the manufacturer provided a critical appraisal of the PRIMA[32] trial using a checklist based on the Centre for Reviews and Dissemination Report.[44] A summary of this critique and the ERG's comments are included in Table 29 in Appendix 1. Overall, the ERG considers that the PRIMA[32] trial was of good design with centralised, random allocation, and safeguards to mitigate against possible bias in monitoring and assessment (the latter are particularly important because the trial was open-label). However, the ERG is concerned about the immaturity of the outcome data reported by the manufacturer; this immaturity is due to the trial's early closure and the, at-present, limited duration of the follow-up data.

The PRIMA[32] trial was an open-label study, thus investigators, sponsor and patients were aware of treatment assignments (after allocation, which was centrally controlled). Blinding is especially important in trials with a primary outcome of PFS as PFS relies on investigator assessment and therefore could be subject to potential assessment bias.[45] However, as the robustness of the PFS assessment by the investigator was corroborated by the results of the Independent Review Committee (IRC) of radiological and clinical data, the manufacturer did not consider this to be an issue in the PRIMA[32] trial; neither did the ERG.

Trial design and statistical considerations

The MS reports that in the initial design stage of the PRIMA[32] trial, an interim analysis was planned after 75% of events (258 out of 344) had been observed. The required number of events was reached in January 2009 after a median follow-up time of 25 months. The results of the interim analysis were reviewed by an independent Data and Safety Monitoring Committee (DSMC) who judged that the PRIMA[32] trial had met its primary objective and recommended premature closure of the trial with all outcomes fully analysed and publicly disclosed.

The ERG notes that the planned interim analysis by the manufacturer was based on the information fraction approach, i.e. after 75% of events had occurred. Although this approach is ideal from a statistical standpoint and frequently practiced, its application in time-to-events monitoring when the events are rare during the study period is debateable since early survival experience with short follow up may not accurately reflect the complete experience with time. One alternative is to use calendar time scale (e.g. yearly, after the first 2 years of recruitment); this is considered to provide a reasonable approximation of information fraction in situations where there are few events during a study period as this can result in longer patient follow up.[46, 47]

The sample size calculations in the PRIMA[32] trial protocol are based on the number of patients required to show a specific difference in median PFS (37.2 months in the observation arm vs 54 months in the RTX arm) between arms. This involves assumptions about the monthly hazard rate in the observation arm (0.0186) and the RTX maintenance therapy arm (0.0186 for 6 months, 0.0121 thereafter). In the latest available PRIMA[32] trial data the 2-year average monthly PFS hazard rate is 9% less than expected in the observation arm, and 30% less in the RTX maintenance arm. Although these early results are clearly encouraging, they indicate that trial events were accruing more slowly than anticipated and that reaching the primary trial objective (to demonstrate a statistically significant difference in median PFS) would require longer follow-up of patients to ensure that these initial indications are not (at least in part) a consequence of random variations.

The ERG notes that the premature closure of this study resulted in a dataset that is currently very immature; at the time of study closure there were 1018 patients enrolled, 505 and 513 in the RTX and observation arms respectively. Of these patients, only 93 (18.4%) in the RTX arm and 174 (33.9%) in the observation arm had experienced an event at the time of analysis. In addition, at the time of closure, none of the patients in PRIMA[32] had been followed up for more than 4 years. The ERG notes that, according to major protocol Amendment 3 (in the CSR), patients should be followed up for 7 years; the fact that none of the patients had been followed up as planned before closure highlights the immaturity of the dataset. [redacted]

[redacted]

The ERG notes that the statistically significant difference in PFS hazard ratio (HR) between RTX and observation was achieved before any difference in median PFS could be estimated. The ERG is concerned that stopping a clinical trial early may lead to an overestimation of treatment benefits and is aware that early evidence of benefit is not always confirmed by later data. The ERG notes that, [redacted]

[redacted]

Protocol amendments

The ERG notes from the MS that five protocol amendments were made to the PRIMA[32] trial. One of these amendments warrants further discussion. Amendment 3 reported in the MS (p.69) states:

> *This amendment modified the primary endpoint of the study from EFS to PFS and increased the total number of study patients from 900 to 1200. The number of PFS events required for the final analysis was also modified, partly to account for a possible six-month lag in rituximab treatment benefit following randomisation. In line with these changes, the number of PFS events that would trigger interim efficacy analyses also increased (from 100 and 150 events for the first and second interim analyses, respectively, to172 and 258 events.*

The change in primary endpoint was made within 6 months of the start of the PRIMA[32] trial. The ERG is of the opinion that the manufacturer changed the primary endpoint in line with the key outcomes of interest as required by the Food and Drug Administration, EMA and NICE.

The utility of PFS as a primary outcome in cancer trials is clearly important, however, the ERG is of the opinion that EFS and time to next lymphoma treatment are also important clinical outcomes. The elements of PFS in the PRIMA[32] trial were: documented disease progression, relapse, or death from any cause. The elements of EFS were: first documented progression, relapse, initiation of a new anti-lymphoma treatment (CTX, radiotherapy, radioimmunotherapy, immunotherapy) or death from

any cause. Assessment of PFS is generally a post-dated event as it will be reliant on the results of laboratory analyses which may take time to reach the clinician from the laboratory. Assessment of EFS is more immediate as it may be determined by the treating physician at the time of the patient's attendance at the clinic rather than waiting on the results of laboratory data. The ERG is of the view that EFS mirrors clinical practice more closely than does PFS and therefore may be informative as a primary trial endpoint; however, the ERG is also aware that as PRIMA[32] is an open-label trial, there is a higher risk of bias in the use of EFS.

Description of the PRIMA trial

The PRIMA[32] trial was conducted in 24 countries worldwide. Large numbers were recruited and randomisation was applied centrally rather than within centre or country. From the CSR (p.48) it appears that suitable monitoring systems were implemented by Roche to ensure uniformity of practice across centres. [redacted]

[redacted]

The ERG considers that results from PRIMA[32] can be generalised to patients in England and Wales.

The majority of patients in the PRIMA[32] trial were treated with R-CHOP induction therapy. In the UK, NICE guidance TA110[27] currently recommends R-CVP as first-line therapy for patients with symptomatic stage III and IV follicular lymphoma. The manufacturer correctly points out that this was consistent with the marketing

authorisation for RTX at the time that the guidance was issued in 2006. However, the licence for RTX has since been broadened by the EMA to include the use of RTX with any CTX. The ERG considers that the R-CTX combinations utilised in PRIMA[32](R-CHOP, R-FCM, R-CVP) are appropriate and reflect those used in clinical practice in the UK. The ERG further notes that the current ESMO[15, 28] and Scottish Medicines Consortium[20] guidelines recommend any combination of CTX with RTX in this indication.

In the PRIMA[32] trial, patient groups were similar at baseline in both the induction and maintenance phases (Table 9). The ERG notes that the data demonstrate the expected gender distribution (similar numbers of males and females) and the usual RCT study population (i.e. slightly younger patient population than might be seen in clinical practice).

Table 9 Patient characteristics: induction and maintenance phases

	Induction phase			Maintenance phase	
Characteristic	**R-CHOP (N=881)**	**R-CVP (N=268)**	**R-FCM (N=44)**	**RTX (N=505)**	**Observation (N=513)**
Male	463 (53%)	137 (51%)	22(50%)	270 (53%)	263 (51%)
Female	418 (47%)	131 (49%)	22 (50%)	235 (47%)	250 (49%)
Age (years) at registration: mean (sd)	55.4 (11.47)	57 (12.66)	51.3 (10.87)	56.5 (11.12)	55.5 (12.06)
Weight in kg mean (sd)	73.3(15.02)	76.0 (15.73)	73.5 (18.92)	74.8 (15.44)	73.3(14.65)
BSA mean (sd)	1.8 (0.21)	1.9 (0.21)	1.8 (0.25)	1.8 (0.21)	1.8 (0.21)
ECOG 0	574 (65%)	161 (60%)	24 (55%)	[redacted]	[redacted]
ECOG 1	272 (31%)	94 (35%)	16 (36%)	[redacted]	[redacted]
ECOG 2	35 (4%)	13 (5%)	4 (9%)	[redacted]	[redacted]
Ann Arbor Stage I	19 (2%)	4 (1%)	0	[redacted]	[redacted]
Ann Arbor Stage II	68 (8%)	22 (8%)	5 (11%)	[redacted]	[redacted]
Ann Arbor Stage III	167 (19%)	53 (20%)	7 (16%)	[redacted]	[redacted]
Ann Arbor Stage IV	627 (71%)	189 (71%)	32 (73%)	[redacted]	[redacted]
FLIPI as reported on the CRF:					
0	33 (4%)	15 (6%)	4 (9%)	[redacted]	[redacted]
1	160 (18%)	39 (15%)	3 (7%)	[redacted]	[redacted]
2	312 (35%)	91 (34%)	20 (45%)	[redacted]	[redacted]
3	242 (28%)	76 (28%)	12 (27%)	[redacted]	[redacted]
4	115 (13%)	38 (14%)	4 (9%)	[redacted]	[redacted]
5	18 (2%)	8 (3%)	1 (2%)	[redacted]	[redacted]

ECOG= Eastern Co-operative Oncology Group; BSA = body surface area; CRF = case report form

Post-progression treatments

In relation to post-progression treatments, the MS reports (p.133) that, at the time of the analyses (Jan 2009), [redacted]

[redacted]

[redacted]

Any post-progression treatment is likely to impact on the outcome of OS. The ERG notes that post-progression treatments described in the MS appear to match those that would be offered in UK clinical practice; however, it is unclear from the data whether the timings of post-progression treatments in the trial would match the timings of post-progression treatments in the UK.

4.3.1 Description and critique of manufacturers outcome selection

The primary outcome reported in the PRIMA[32] trial is PFS. The secondary outcomes are EFS, OS, time to next anti-lymphoma treatment, time to next CTX treatment, overall response rate (RR) at the end of maintenance/observation, transformation rate at first progression, QoL and safety. The ERG considers that the outcomes selected by the manufacturer are appropriate to this disease area and match those stated in the final scope issued by NICE;[14] however, the dataset reporting these outcomes is too immature to meaningfully inform clinical decision-making for patients with fNHL.

4.3.2 Describe and critique the statistical approach used

The MS identified only one appropriate study (PRIMA[32]) and so no meta-analyses or indirect comparisons were undertaken.

The PRIMA[32] trial was originally designed with EFS as the primary outcome, but as noted above, this was changed to PFS within 6 months of the trial starting.

Randomisation in the PRIMA[32] trial was stratified according to induction regimen, centre, region and response to induction treatment (CR/CRu or PR) and was carried out centrally. The ERG considers this to be appropriate.

The PRIMA[32] trial was stopped early on the recommendation of the DSMC. In the submitted dataset, there is limited follow-up; this means that a very small number of

patients have experienced an event and the manufacturer has been unable in most instances, to calculate the median time to event (MTE).

The primary outcome was PFS based on the investigator assessment (Jan 2009). Progression-free survival was also assessed by the IRC. Data were available for 887 patients, of these, [redacted] from the RTX and observation arms respectively were reported to have experienced an event (i.e. disease progression or death). However, in the MS the percentage of patients experiencing an event is calculated using the total number of patients in each arm rather than the total number of patients who have been assessed by the IRC. This leads to slightly different numbers [redacted]% in the RTX arm and [redacted] in the observation arm.

Several subgroup analyses were undertaken to assess the potential impact of age (≥ 60 years, < 60 years), gender (male, female), pre-induction FLIPI score (≤ 1, 2, ≥ 3), induction treatment (R-CHOP, R-CVP, R-FCM) and response to induction treatment (CR/CRu, PR) on the treatment effect. The ERG considers these subgroup analyses to be acceptable as they were all pre-specified. However, the manufacturer states that the PRIMA[32] trial was not designed to show differences between subgroups and the ERG is of the opinion that since the full dataset has not yet reached maturity, results from any subgroup analyses hold little meaning.

The MS describes a number of protocol violations in the PRIMA[32] trial: 23 patients (14 patients in the RTX arm and 9 in the observation arm) started a new anti-lymphoma treatment before documented progression; 16 patients (10 patients in the RTX arm and 6 in the observation arm) received RTX as part of their new treatment; and 3 patients in the RTX arm received additional cycles of RTX through investigator error.

The MS clearly describes the patient withdrawals from the PRIMA[32] trial. More patients in the observation arm than in the RTX arm withdrew from the study (162 patients vs 101 patients; 32% vs 20%).

4.4 Results

4.4.1 Induction phase

[redacted]1193 patients) [redacted]

[redacted](75%);_[redacted]

[redacted]

[redacted]1193)[redacted]

[redacted]1193). The response rates for the induction phase are described in Table 10. The majority of patients received induction with R-CHOP. The overall RR is broadly similar across the groups.

Table 10 Response to induction (PRIMA trial)

Response	**R-CHOP (N=881) n(%)**	**R-CVP (N=268) n(%)**	**R-FCM (N=44) n(%)**
All responders[a]	[redacted]	[redacted]	[redacted]
Non-responders	[redacted]	[redacted]	[redacted]
95% CI for response rates[b]	[redacted]	[redacted]	[redacted]
Differences in response rates (95% CI)[c]		[redacted]	[redacted]
CR/CRu	[redacted]	[redacted]	[redacted]
95% CI for CR/CRu rates[b]	[redacted]	[redacted]	[redacted]
Differences in CR/CRu response rates (95% CI)[c]		[redacted]	[redacted]
PR	[redacted]	[redacted]	[redacted]
95% CI for PR rates[b]	[redacted]	[redacted]	[redacted]
Difference in PR rates (95% CI)[c]		[redacted]	[redacted]
SD	[redacted]	[redacted]	[redacted]
95% CI for SD rates[b]	[redacted]	[redacted]	[redacted]
PD	[redacted]	[redacted]	[redacted]
95% CI for PD rates[b]	[redacted]	[redacted]	[redacted]
Not evaluated (NE)	[redacted]	[redacted]	[redacted]
95% CI for NE rates[b]	[redacted]	[redacted]	[redacted]
Missing (no response assessment)	[redacted]	[redacted]	[redacted]

a Patients with a response of CR, CRu or PR at the end of the induction treatment

b 95% CI for one sample binomial using Pearson-Clopper

c Approximate 95% CI for difference of two rates using Hauck-Anderson method

4.4.2 Maintenance/observation phase

The MS presents a series of analyses describing the maintenance/observation phase data. The analyses are extensive, spanning from p.98 of the MS to p.148. Data from two time points are described in the clinical-effectiveness section of the MS. The main data are derived from a cut-off point of January 2009 with supplementary data taken from an analysis in January 2010. In the clarification response, the manufacturer has also provided further data from January 2010 and a 'snapshot' of PFS data taken at June 2010.

Other analyses that are secondary or confirmatory are summarised in Table 31 in Appendix 2. The secondary analyses were carried out to confirm the robustness of the PFS outcome finding and include PFS for the per protocol (PP) populations as assessed by investigators and the IRC and comparisons between investigator assessments and IRC assessments. In summary, the analyses appear to support the investigator assessed PFS results observed in the intention to treat (ITT) population.

The comparison between investigator and IRC assessments demonstrated a high concordance rate [redacted] in terms of progression in both arms; but for the assessment of the timing of disease progression, a lower concordance rate was noted [redacted]) in both arms. It is stated in the MS that discordance in timing of disease progression was mostly due to events that occurred earlier based on IRC assessments compared with investigator assessments (p.115). The ERG notes from the CSR (p.41) [redacted]. The manufacturer notes that IRC reviews were not conducted for scans collected after the first cut-off date of January 2009.

Primary endpoint: January 2009

The ERG reiterates that it considers the results reported in this section are derived from a very immature dataset. This is clearly illustrated in Table 11, which summarises the outcomes for the maintenance phase of the PRIMA[32] trial up to January 2009. The median follow-up was 25 months. For the primary outcome of investigator-assessed PFS, the MTE is not estimable and it appears that only 25% of the trial population had experienced an event (disease progression, relapse, or death) within the time frame of the data analysis. The ERG notes that the investigator-assessed PFS is greater (by 10 months) than the IRC-assessed PFS on the 25th percentile; this may be explained by the fact that in the IRC assessment there were

missing scans and the IRC occasionally issued earlier diagnoses of disease progression than the investigators.

The data in Table 11 demonstrate a statistically significant reduction in risk of disease progression between the RTX and observations arms in favour of RTX (HR=0.50; 95% CI 0.39 to 0.64; $p < 0.0001$). According to Kaplan-Meier estimates, 89% of patients in the RTX arm and 82% of patients in the observation arm were progression free (investigator assessed) at 1 year, whilst at 2 years, 82% of patients in the RTX arm and 66% of patients in the observation arm were progression free (investigator assessed).

Table 11 PRIMA trial primary outcome PFS (Jan 2009)

Endpoint	RTX (N=505)	Observation (N=513)	HR (95% CI)	p-value
Primary: Investigator-Assessed PFS				
Median time to event	NE	NE		
25th percentile	1096 days (36.0 months)	507 days (16.7 months)	HR = 0.50 (0.39 to 0.64)	p < 0.0001
1- year PFS rate (95% CI)	0.89 (0.87 to 0.92)	0.82 (0.79 to 0.85)		
2- year PFS rate (95% CI)	0.82 (0.79 to 0.86)	0.66 (0.62 to 0.71)		
Primary: IRC-Assessed PFS				
Median time to event	[redacted]	[redacted]		
25th percentile	[redacted]	[redacted]	HR = 0.54 (0.42 to 0.70)	p < 0.0001
1-year PFS rate (95% CI)	[redacted]	[redacted]		

NE = not estimable
p-values and hazard ratios were calculated using the stratified log-rank test and stratified Cox regression for time-to-event endpoints, respectively. Stratification factors were induction treatment received and response to induction treatment. p-values for response rate were calculated using the χ^2 test

January 2010

The outcome data for PFS up to January 2010 are described in Table 12. The median follow-up is 36 months. The ERG notes that the MTE in the RTX arm remains not estimable, whilst the MTE in the observation arm is reported.

At 3 years, [redacted]

Table 12 PRIMA trial primary outcome PFS (Jan 2010)

Endpoint	RTX (N=505)	Observation (N=513)	HR (95% CI)	p-value
Primary: Investigator-Assessed PFS				
Median time to event	[redacted]	[redacted]		
25th percentile	[redacted]	[redacted]	[redacted]	[redacted]
3-year PFS rate (95% CI)	[redacted]	[redacted]		

June 2010

In the clarification response, the manufacturer provided PFS data up to June 2010. This is presented in Table 13 which summarises the PFS outcomes from each data cut-off point. The ERG notes that after 38 months of follow-up, too few patients had experienced an event to allow calculation of the median PFS for the RTX arm of the PRIMA[32] trial. The ERG considers that before the clinical efficacy of a maintenance treatment with RTX can be properly assessed a more mature dataset, based on longer follow-up and including more events, is required.

Table 13 Summary of PFS from PRIMA from January 2009 to June 2010

PRIMA clinical data cut off date	Median follow-up	RTX Median PFS N=505	Observation Median PFS N=513	HR (95% CI)	p-value*
14th January 2009	25 months	NE	NE	0.50 (0.39 to 0.64)	p < 0.0001
15th January 2010	36 months	NE	[redacted]	[redacted]	[redacted]
14th June 2010	38 months	NE	[redacted]	[redacted]	[redacted]

* p-values and hazard ratios were calculated using the stratified log-rank test and stratified Cox regression for time-to-event endpoints, respectively. Stratification factors were induction treatment received and response to induction treatment. p-values for response rate were calculated using the χ^2 test. NS= not specified.

Secondary outcomes: January 2009

The data for the secondary outcomes at January 2009 are described in Table 14. The ERG notes that the manufacturer was unable to calculate the MTE for the RTX arm on any outcome and was only able to calculate MTE for EFS for the observation arm.

[redacted]

[redacted]

[redacted]time to next anti-lymphoma treatment (HR=0.61; CI 0.46 to

0.82; p = 0.0003), and time to next CTX treatment (HR=0.60; CI 0.44 to 0.82; p = 0.0011) are reported. The manufacturer notes that [redacted]

[redacted]

The overall RR at the end of the maintenance/observation phase (based on 787 patients) was statistically significantly greater in the RTX arm than in the observation arm (74% vs 55%, p < 0.0001). Additionally, the proportion of patients with a CR was greater in the RTX arm (66.8%) than in the observation arm (47.7%) and there was a greater proportion of patients in the observation arm (40.7%) with progressive disease than in the RTX arm (20.3%).

Table 14 PRIMA trial secondary outcomes January 2009

Endpoint	RTX (N=515)	Observation (N=505)	HR/OR (95% CI)	p-value
Secondary Endpoints				
Event-free survival				
Median time to event	[redacted]	[redacted]		
25th percentile	[redacted]	[redacted]	[redacted]	[redacted]
One-year event-free rate (95% CI)	[redacted]	[redacted]		
Overall survival				
Median time to event	[redacted]	[redacted]		
25th percentile	[redacted]	[redacted]	[redacted]	[redacted]
One-year event-free rate (95% CI)	[redacted]	[redacted]		
Time to next anti-lymphoma treatment				
Median time to event	NE	NE		
25th percentile	1135 days (37.3 months)	746 days (24.5 months)	HR = 0.61 (0.46 to 0.80)	p = 0.0003
One-year event-free rate (95% CI)	0.92 (0.89 to 0.94)	0.89 (0.87 to 0.92)		
Time to next CTX treatment				
Median time to event	NE	NE		
25th percentile	1135 days (37.3 months)	884 days (29.0 months)	HR = 0.60 (0.44 to 0.82)	p = 0.0011
One-year event-free rate (95% CI)	0.92 (0.90 to 0.95)	0.91 (0.89 to 0.94)		
Overall response rate at the end of maintenance/observation				
N (excluding patients still ongoing maintenance)	N = 389	N = 398		
Responders (CR, CRu, PR)	288 (74%)	219 (55%)	Diff: 19.01 (12.3 to 25.7)	p < 0.0001
Non-responders	101 (26%)	179 (45%)	OR = 2.33 (1.73 to 3.15)	
Patients with complete response (CR/CRu)	260 (66.8%)	190 (47.7%)		
partial response (PR)	28 (7.2%)	29 (7.3%)		
stable disease (SD)	0 (0%)	1 (0.3%)		
progressive disease (PD)	79 (20.3%)	162 (40.7%)		
Transformation Rate at first progression				
Patients with progression	[redacted]	[redacted]		
Transformation	[redacted]	[redacted]	[redacted]	[redacted]
No transformation (no progression/missing)	[redacted]	[redacted]	[redacted]	

Dif.: difference in rates; NE: not estimable

p-values and hazard ratios were calculated using the stratified log-rank test and stratified Cox regression for time-to-event endpoints, respectively. Stratification factors were induction treatment received and response to induction treatment. p-values for response rate were calculated using the $\chi 2$ test, and odds ratios were calculated by using logistic regression (response rate analyses were unadjusted)

Secondary outcomes: January 2010

The data for the secondary outcomes up to January 2010 are described in Table 15. The ERG notes the lack of any MTE calculations for the RTX arm. [redacted] The manufacturer notes that [redacted]

The overall RR at the end of the maintenance/observation phase (based on 1018 patients) was statistically significantly greater in the RTX arm than in the observation arm (78% vs 60%; $p < 0.0001$). Additionally, the proportion of patients with a CR was greater in the RTX arm (71.5%) than in the observation arm (52.2%) and there was a greater proportion of patients in the observation arm (35%) with progressive disease than in the RTX arm (17%).

Table 15 PRIMA trial secondary outcomes January 2010

Endpoint	RTX (N=515)	Observation (N=505)	HR/OR (95% CI)	p-value
Secondary Endpoints				
Event-free survival				
Median time to event	[redacted]	[redacted]		
25th percentile	[redacted]	[redacted]	[redacted]	[redacted]
One-year event-free rate (95% CI)	[redacted]	[redacted]		
Overall survival				
Median time to event	[redacted]	[redacted]		
25th percentile	[redacted]	[redacted]	[redacted]	[redacted]
One-year event-free rate (95% CI)	[redacted]	[redacted]		
Time to next anti-lymphoma treatment				
Median time to event	[redacted]	[redacted]		
25th percentile	[redacted]	[redacted]	[redacted]	[redacted]
One-year event-free rate (95% CI)	[redacted]	[redacted]		
Time to next CTX treatment				
Median time to event	[redacted]	[redacted]		
25th percentile	[redacted]	[redacted]	[redacted]	[redacted]
One-year event-free rate (95% CI)	[redacted]	[redacted]		
Overall Response Rate at the end of maintenance/observation				
N (excluding patients still ongoing maintenance)	[redacted]	[redacted]		
Responders (CR, CRu, PR)	[redacted]	[redacted]	[redacted]	[redacted]
Non-responders	[redacted]	[redacted]	[redacted]	
Patients with complete response (CR/CRu)	[redacted]	[redacted]		
partial response (PR)	[redacted]	[redacted]		
stable disease (SD)	[redacted]	[redacted]		
progressive disease (PD)	[redacted]	[redacted]		
Patients with progression	[redacted]	[redacted]		
Death without progression	[redacted]	[redacted]		

*the ERG notes that this NE is likely to be a typing error as this figure was reported at an earlier time-point

[redacted]

The manufacturer's clarification response also documents that at the end of the maintenance/observation phase [redacted] patients evaluated in the RTX maintenance arm were [redacted] patients evaluated at the same time in the observation arm [redacted]

[redacted]

Subgroup analyses

Pre-specified subgroup analyses were carried out on the following factors: age (< 60 years or > 60 years), gender (male or female), FLIPI index (≤1, 2 or ≥3), type of induction CTX and response to induction therapy (CR/CRu or PR). The manufacturer states that the effect of the study treatment on patients was similar across all groups and that all patients benefited from RTX maintenance therapy 'regardless of the examined patient characteristics.'(MS, p.107) This was noted at the 2010 cut-off (MS, p.147) with the risk of disease progression or death reduced in the RTX arm compared to the observation arm in all of the subgroups tested ([redacted].

The ERG considers that the results of any subgroup analyses should be regarded with caution, given that lack of events recorded in the main dataset has precluded any meaningful conclusions being drawn from the data available. The manufacturer has also stated that the PRIMA[32] study was not powered to show significant differences between subgroups (MS, p.33).

4.5 Health-related quality of life

Health-related QoL was assessed in the PRIMA[32] trial using the FACT-G[43] and EORTC QLQ-C30[42] questionnaires. The FACT-G[43] is a 27-item compilation of general questions divided into four primary QOL domains: Physical Well-Being, Social/Family Well-Being, Emotional Well-Being, and Functional Well-Being. It is considered appropriate for use with patients with any form of cancer. The maximum score on FACT-G[43] is 112; a high score indicates a better QoL.

The EORTC QLQ-C30[42] is a questionnaire developed to assess the quality of life of cancer patients and is commonly utilised in clinical trials. It consists of 30 items relating to the categories of global health status/QoL, functional scale and symptom

scale. Better quality of life is indicated by high scores (between 0 and 100) on the global health/QoL and functional scales and by a lower score on the symptom scale.

The MS states that the questionnaires were distributed to patients at screening, at the end of the induction phase, after one year of the maintenance/observation phase, at the end of the maintenance/observation phase and every year for the 5 year follow-up period. Questionnaires were not distributed once a patient had progressed.

The data presented for both questionnaires span from screening to up to 2 years beyond the maintenance/observation phase. Similar numbers of patients from both arms completed questionnaires at each time point. However, the ERG notes that the majority of the data were collected at screening and after induction, with more than 300 patients from each arm completing forms. The number of patients completing forms after 1 and 2 years of maintenance/observation is reduced (as a result of progressive disease). As yet, there are few respondents (less than six) at 1 and 2 years following the end of maintenance/observation. The MS does not indicate whether or not the completion rates to both questionnaires were considered by the manufacturer to be satisfactory.

Results of the patient responses to the FACT-G[43] questionnaire are reported by the manufacturer [redacted] Table 39 in the MS, labelled as 'Summary of FACT-G scores over time' does not contain any data. However, scores for each domain of the questionnaire for each trial arm across time are presented in the CSR.[32] [redacted]

Results of the patient responses to the EORTC QLQ-C30[42] are also reported to be similar at baseline across the two arms of the trial and did not change substantially over time. These data are presented in Table 40 of the MS (MS, p.142). The scores for each subscale are presented in the CSR[32] and the ERG [redacted]

[redacted]

4.6 Safety/adverse events

The manufacturer presents extensive and very detailed records of safety data assessed in the PRIMA[32] trial; the AE section of the MS spans from pages 155 to 209. This section of the ERG report is a synopsis of the information presented in the MS. Table 23 in Appendix 3 provides a summary table of the manufacturer's overview of each safety analysis for the maintenance phase of the PRIMA[32] trial. For more detailed information, the reader is referred to the MS.

4.6.1 Induction phase

The MS provides an overview of the toxicities and serious adverse events (SAEs) recorded during the induction phase of the PRIMA[32] trial. Table 16 replicates the data reported (MS, p.154). The manufacturer reports that the safety profile of RTX in combination with the CHOP, CVP and FCM was consistent with the known safety profile of these induction regimens. There were no new or unexpected safety findings. Further detail relating to the toxicities and AEs recorded during the induction phase of the PRIMA[32] trial is reported in pages 154 to 160 of the MS.

Table 16 Toxicities and adverse events from the PRIMA trial (induction)

Safety parameter	**R-CHOP N = 881 n(%)**	**R-CVP N = 268 n(%)**	**R-FCM N = 44 n(%)**
Toxicities [a]	[redacted]	[redacted]	[redacted]
Adverse events[b]	[redacted]	[redacted]	[redacted]
Serious adverse events	[redacted]	[redacted]	[redacted]
Grade 3/4 AEs	[redacted]	[redacted]	[redacted]
Grade 5 (fatal) AEs	[redacted]	[redacted]	[redacted]
Related AEs	[redacted]	[redacted]	[redacted]

[redacted]

4.6.2 Maintenance/observation phase

The MS reports that 1009/1019 patients received at least one dose of RTX maintenance or attended one observation visit during the maintenance/observation phase and were included in the maintenance safety analysis population. Of these, 501 patients were in the RTX arm and 508 were in the observation arm.

In the MS, it is acknowledged that Grade 1 infections and Grade 1 and 2 AEs other than infections were not recorded during the maintenance/observation phase except as toxicities on the checklist toxicity case report form (CRF). It is also reported that mandatory collection of SAE data for the observation arm were not clearly specified in the study protocol: an error that was addressed shortly before data cut-off in January 2009. However, the manufacturer has assessed the proportion of AEs that were reported as SAEs in both arms of the PRIMA[32] trial and is confident that systematic under-reporting of SAEs in the observation arm did not occur.

The MS reports on the toxicities and AEs recorded during the maintenance/observation phase of the PRIMA[32] trial. These are summarised in Table 17. The manufacturer notes that the incidence of AEs was higher in the RTX arm compared with the observation arm. The difference was mainly due to infections and infestations (37% of patients in the RTX arm vs 22% of patients in the observation arm). However, the manufacturer also states that 'RTX maintenance therapy was well tolerated and no unexpected safety findings were observed' (MS, p.162). The clinical advisor to the ERG agrees that clinical experience shows RTX to be well-tolerated by patients.

Table 17 Toxicities and AEs from PRIMA 2009 (maintenance/observation)

Safety parameter	RTX N = 501 n (%)	Observation N = 508 n (%)
Toxicities[a]	485 (97)	459 (90)
Adverse events[b]	263 (52)	179 (35)
Grade 3/4 AEs	114 (23)	81 (16)
Serious adverse events	95 (19)	63 (12)
Withdrawal from treatment due to toxicity	10 (2)	1 (<1)
AEs leading to treatment discontinuation	19 (4)	8 (2)
AEs leading to dose modification	30 (6)	–
AEs leading to death	3 (<1)	2 (<1)
Infection AEs (Grade ≥ 2)	184 (37)	114 (22)
Grade 3/4 infections	22 (4)	5 (<1)
AEs occurring within one day after treatment/observation visit	61 (12)	46 (9)
Total deaths	13 (3)	18 (4)
Death due to cause other than lymphoma	3 (<1)	6 (1)

a Toxicities are based on the checklist CRF page (regardless of grade).
b Includes Grade 3–5 toxicities, Grade 2–5 infections, and SAEs regardless of grade, as recorded on the AE CRF pages.

The manufacturer reports that analysis of data from the later cut-off point (January 2010) revealed a slight increase in the number of AEs and SAEs recorded since January 2009; however, no change in the safety profile was observed. Further information provided by the manufacturer in response to a clarification request from NICE is summarised in Table 18. These data are used by the manufacturer to support its economic case.

Table 18 Toxicities and AEs from PRIMA 2010 (maintenance/observation)

Safety Parameter	RTX N = 501 n (%)	Observation N = 508 n (%)
Adverse events[b]	[redacted]	[redacted]
Grade 3/4 AEs	[redacted]	[redacted]
Serious adverse events	[redacted]	[redacted]
AEs leading to treatment discontinuation	[redacted]	[redacted]
Toxic death	[redacted]	[redacted]

4.7 Summary of submitted evidence

The MS cites evidence from a single, unpublished, open-label RCT, the PRIMA[32] trial; no meta-analyses were therefore reported. The PRIMA[32] trial was a well-designed RCT. Patients with previously untreated fNHL, who had responded to R-CTX induction therapy (n=1018) were randomised to receive either RTX maintenance therapy or observation for 2 years or until progression. The trial was stopped early following an interim analysis after 75% of the planned events had

occurred. To date, the available PRIMA[32] data appear to demonstrate that for primary and secondary endpoints, patients in the RTX arm have improved outcomes compared to patients in the observation arm. There did not appear to be any differences between the RTX or observation arms on the outcome of QoL. No unexpected AEs were noted, however, patients in the RTX arm experienced more infections than patients in the observation arm.

4.8 Summary of results

4.8.1 Clinical results

- The main source of clinical evidence described in the MS is from the unpublished PRIMA[32] trial
- The PRIMA[32] trial includes patients with untreated fNHL (90% stage III/IV) who responded (CR/CRu/PR) to R-CTX induction therapy and then received either RTX monotherapy or observations for 2 years or until disease progression
- [redacted]
- On the secondary endpoints of EFS, time to next anti-lymphoma treatment, time to next CTX treatment and overall RR, [redacted] of RTX at the January 2010 data cut-off. However, the only calculable MTE was for EFS in the observation arm
- For the outcomes of OS and transformation rate at first progression, no meaningful conclusions could be drawn due to lack of events
- Subgroup analyses that took into account age, gender, FLIPI prognosis, type of induction CTX and response to induction indicated that all patients benefitted equally from RTX treatment. However, the trial was not powered to detect differences in subgroups
- The incidences of AEs were higher in the RTX arm compared with the observation arm; however, no unexpected safety findings were noted
- No statistically significant differences were reported on the outcome of QoL as measured by FACT-G[43] and EORTC-QLQ C30[42] questionnaires.

4.8.2 Clinical issues

- The ERG is concerned that the dataset presented is immature. Lengthier follow-up with more events is required before informed decisions can be made regarding the clinical effectiveness of RTX as a first-line maintenance therapy compared to observation
- Follicular NHL is a treatable disease, but it is not curable. Patients experience recurring and remitting disease over many years. Each successive relapse becomes more difficult to treat. It is not clear if there are optimal time points for the use of RTX treatment.

5 ECONOMIC EVALUATION

5.1 Introduction

This section provides a structured critique of the economic evidence submitted by the manufacturer of RTX. The two key components of the MS are (i) a systematic review of the relevant literature and (ii) a report of the manufacturer's *de novo* economic evaluation. See Table 19 for a summary of key information points. The manufacturer also provided an electronic version of the EXCEL based economic model.

Table 19 Key information in the MS

Key information	Pages (MS)	Key tables/figures (MS)
Details of the systematic review of the economic literature	231-233	
Model structure	236-241	Figures 28-30; Table 88
Technology	241-242	
Clinical parameters and variables	243-266	Figures 31-42; Tables 90-98
Measurement and valuation of health effects and adverse events	267-277	Figure 43; Tables 100-102
Resource identification, valuation and measurement	277-290	Tables 103-109
Sensitivity analysis	291-293	Figure 46; Table 110
Results	295-311	Figures 48-53; Table 111-126
Validation	Clarification response (2)	
Subgroup analysis	Not applicable	
Strengths and weaknesses of economic evaluation	314-316	
Assessment of factors relevant to other parties	317-321	Table 127-130

5.2 Overview of manufacturer's cost-effectiveness review

The MS provides a description of the review of published cost-effectiveness evidence undertaken by the manufacturer. The aim of the review was to identify cost-effectiveness studies of RTX as a first-line maintenance treatment in fNHL. The databases searched and the search terms used appear to be reasonable; however, only exclusion criteria were explicitly stated. The search by the manufacturer did not identify any relevant studies for inclusion in the review. Although there is no mention of searching in-house databases for relevant studies, the ERG is confident that no relevant published studies are available for inclusion in the review.

5.3 Overview of manufacturer's economic evaluation

The manufacturer undertook a *de novo* economic evaluation of RTX first-line maintenance therapy (intervention) compared with observation (comparator) for the treatment of patients with fNHL responding to first-line induction. The economic model was developed over a 25 year life-time time horizon in order to capture the lifetime costs and quality adjusted life years (QALYs) of an average patient with fNHL from the perspective of the NHS and Personal Social Services (PSS). In the model, the manufacturer assumed that the average age of the cohort was 55.5 years, body weight was 74.04kg, height was 168.54cm and body surface area (BSA) was 1.835m^2.

5.3.1 Description of the manufacturer's economic model

In the model patients are assumed to be in one of four possible discrete health states at any given time: "progression free survival/first-line maintenance" (PFS1), "progression free survival/second-line treatment" (PFS2), "progression" (PD) or "death". The structure of the economic model is shown in Figure 2.

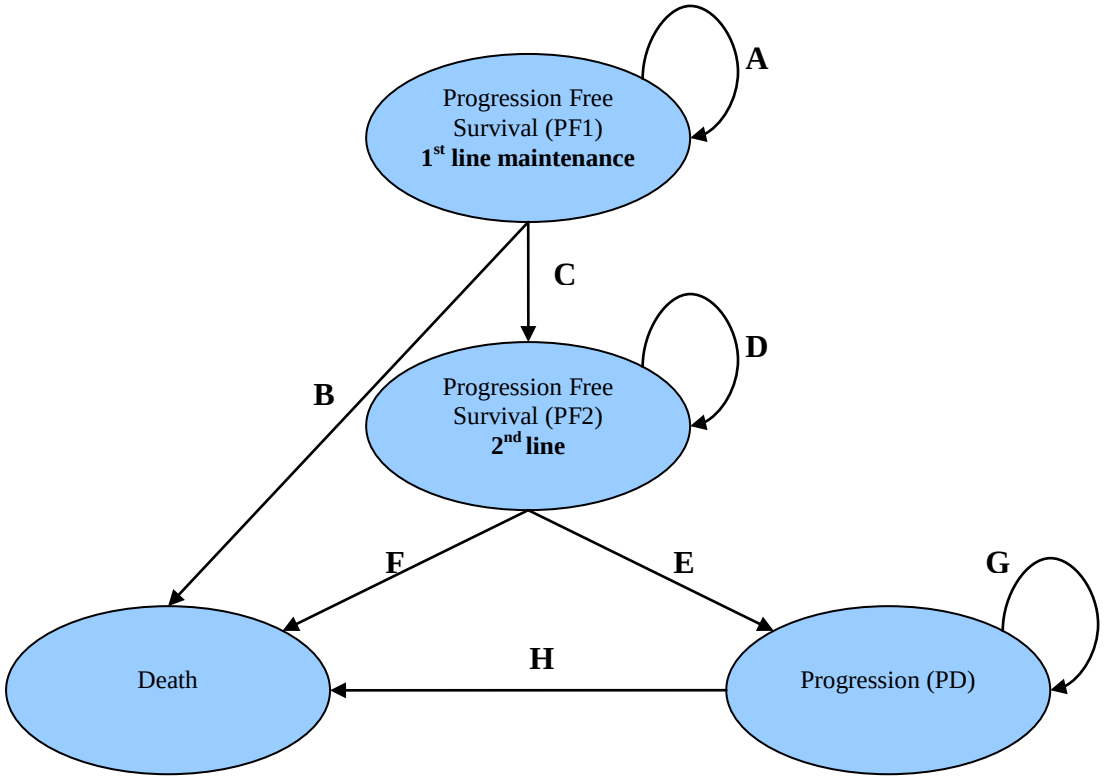

Figure 2 Structure of the manufacturer's model

The economic model developed by the manufacturer is a four state Markov model with a cycle length of one month. Patients enter the model in the PFS1 state having successfully completed induction therapy with R-CTX. At the end of each cycle patients remain in PFS1 or move to PFS2 or die. Once a patient is in the PFS2 health state, a patient may remain in this health state, die at the end of each cycle or move to PD. Patients in PD cannot move back to PFS2 within the model, patients can either remain in PD or die at the end of each cycle. Death is an absorbing health state within the model.

5.3.2 Parameters and values

The base case model parameters and values used in the submitted economic model and described in the MS are presented in Table 20. For additional details of unit costs used in the model please see Table 35 in Appendix 7.

Table 20 Parameters and values used by the manufacturer in the economic model

Model variable	Value	Source
Costs associated with rituximab		
Technology cost (rituximab)	£174.63 (100mg vial) £873.13 (500mg vial)	BNF 56
Mean cost (rituximab)*	£1,222.39 per cycle	BNF 56
Administration cost	£251.4 per infusion	NHS Ref Cost Schedule 08-09
Pharmacy costs	£32 per infusion	PSSRU 2009
Drug costs (PFS2)		
R-CHOP induction	£11,308 (total)	BNF 56
R-CVP induction	£10,284 (total)	BNF 56
R-maintenance	£1,222.39 per cycle	BNF 56
CHOP induction	£1,529 (total)	BNF 56
CVP induction	£504 (total)	BNF 56
Supportive care costs		
PFS1	£42.08 per month	NHS Ref Cost Schedule and assumption
PFS2	£137.04 per month	
PD (cost includes salvage)	£500.53 per month	NICE 2004 /BNF 57/Van Oers
Adverse event costs		
Average cost of treating AE (rituximab)	£272.60	
Average cost of treating AE (observation)	£59.63	
Neutropenia	£3,272 per episode	Not stated
Depression	£44 per episode	Not stated
Pneumonia	£2,494 per episode	Not stated
Dysrhythmias	£606 per episode	Not stated
Infection	£1,077 per episode	Not stated
Granulocytes	£1,514 per episode	Not stated
Deep vein thrombosis	£792 per episode	Not stated
Hazard ratio		
Rituximab vs observation	0.55 (95%CI: 0.44 to 0.68)	PRIMA (snapshot June 2010)
Utility values		
PFS1	0.88 (95%CI: 0.81 to 0.95)	Pettengell/Wild
PFS2	0.79 (95%CI: 0.72 to 0.86)	Pettengell/Wild
PD	0.62 (95%CI: 0.48 to 0.76)	Pettengell/Wild

*Assuming BSA=1.84m^2

5.3.3 Treatment effectiveness within the MS

Clinical data from a variety of sources were utilised to populate the economic model. The primary source of clinical data was the PRIMA[32] study. As the PRIMA[32] data were not mature enough to show a difference in OS at the time of analysis, data from the EORTC 20891[29, 30] trial were also used to inform the economic evaluation.

PRIMA study

The economic evaluation is based on the latest PRIMA[32] clinical dataset (median follow up of 38 months) using data from the 14th June 2010 snapshot. The Kaplan-Meier estimated median PFS time was based on the median time for all patients, censoring those not experiencing an event. The HR for PFS (RTX vs observation) was estimated at [redacted] The PFS trial data were extrapolated using a parametric function for the two arms of the study deriving transition probabilities per cycle for patients in the observation and RTX maintenance states. The rate of death observed while patients were progression free in the PRIMA[32] trial was used to estimate the monthly probability of dying in PFS1. Data from PRIMA[32] demonstrating the percentage of patients that relapse within one year of stopping treatment with RTX were used to determine the appropriate second-line therapy assumption for each patient group.

EORTC 20981 study

The economic evaluation employs the latest EORTC 20981[30] clinical dataset (median follow up of 6 years) using data from the Van Oers et al[30] report published in June 2010. These data were used to determine the probabilities of progression to PD for patients that (i) receive R-CTX induction followed by RTX maintenance or (ii) receive CTX induction. Individual treatment transition probabilities to identify which patients die while in PFS2 were derived using the rate of death while in PFS from the EORTC 20981[30] trial. Although not captured directly by the trial, EORTC[30] data were used to determine the probability of dying in PD; PFS and OS data were combined to estimate post-progression survival. To estimate post-progression treatments, AEs and their associated costs, the manufacturer had to use EORTC data from the 2006 cut-off[29] (49 month median survival) as these data were the latest available to Roche at the time of writing.

5.3.4 Population

The economic evaluation is based primarily on the clinical-effectiveness results of the PRIMA[32] and the EORTC 20981[29, 30] trials. Therefore the population in the economic evaluation reflects the patients enrolled and randomised in the (first-line) induction and maintenance phase of the PRIMA[32] trial. The model also used efficacy data from the patients enrolled and randomised in the (second-line) induction and maintenance phase of the EORTC 20981[29, 30] trial. Whether the patients in PRIMA[32] are sufficiently similar to the patients in the EORTC 20981[29, 30] trial has already been debated in this ERG report (see section 3.5).

5.3.5 Comparator technology

The comparator technology to RTX maintenance in the economic evaluation is observation only (otherwise known as watch and wait therapy or no treatment) in responding patients after first-line treatment until relapse. The manufacturer does not consider Zevalin as a comparator in the economic evaluation.

5.3.6 Health related quality of life

In the PRIMA[32] trial, EQ-5D questionnaires were not routinely administered to patients; HRQoL data were collected in the form of the QLQ-30 [42] and FACT-G [43] questionnaires. The manufacturer reports that there were no differences observed in HRQoL between patients in the RTX arm and patients in the observation arm. A systematic review was conducted by the manufacturer in order to identify HRQoL studies relevant to RTX as a maintenance treatment for patients in fNHL. Three reports[48, 49 50] of a single study were identified by the manufacturer as being relevant to the decision problem. The ERG notes that the by Oxford Outcomes study [49] had been originally commissioned by Roche Products Ltd to elicit utilities for health states associated with fNHL as part of a multi-centre patient based study on QoL in fNHL. In the MS the manufacturer uses the following utility values from the Oxford Outcomes:[49] PFS1= 0.88 (disease free); PFS2 = 0.79 (remission/full response); PD= 0.618.

5.3.7 Resources and costs

In the economic model the manufacturer identifies four major types of costs: NHS costs; intervention and comparators' costs, health state costs and AE costs. Main sources of cost data are the NHS Reference Cost Schedule 08-09[51] and the BNF (56-59).[52]

NHS costs: Cost of administering RTX

The manufacturer assumed that RTX was administered as a hospital day case procedure and used the NHS Reference Cost Schedule[51] (SB15Z) as the source; it was also assumed that it would take one hour to make up the RTX infusion and this additional cost was taken from the Unit Costs of Health and Social Care costs (PSSRU 2009[53]).

NHS costs: Supportive care costs

The manufacturer commissioned a study to identify the resources involved in clinical practice for treating patients with fNHL. The results of an expert consultation demonstrated that clinical practice is variable across treatment centres but consistency was found in the following core aspects: (i) current practice is observation/watch and wait for patients who respond to first-line induction therapy (ii) clinical experts considered that resource use (in supportive care) will not change as a result of first-line maintenance with RTX.

The average cost of supportive care during maintenance in PFS1 was calculated for three phases (1-2 years; 3-4 years and 5 years); actual costs and sources are reported in the MS (MS, Tables 103 and 104). The average supportive care cost in PFS1 was estimated to be the same in the RTX and observation arms (£42.08 per month).

The average cost of supportive care during PFS2 was calculated based on an assumed average duration of 60 months (4.5 months of second-line induction therapy; 24 months of maintenance therapy and 30 months of remission); actual costs and sources are reported in the MS (MS, Table 105). The average supportive care cost in PFS2 was estimated to be the same in the RTX and observation arms (£137.04).

Using data from two distinct sources (NICE 2004[54]; post-progression treatment patient level data from the EORTC 20891[29, 30] study), the manufacturer estimated the total average monthly supportive care cost in PD for both groups of patients to be £500.53.

Intervention and comparator' costs

The manufacturer provides a table (MS, Table 107) which summarises the unit costs associated with RTX as used in the economic model. The MS presents the total drug costs in PFS2 and states that they are taken from BNF 56 but the text of the MS does not offer any further detail on the methods used to estimate costs; the details are only available from the economic model.

Health state costs

The manufacturer provides a table (MS, Table 108) which lists all of the health states and associated costs used in the economic model.

Adverse event costs

The manufacturer provides a table (MS, Table 109) which lists the main AEs and the costs of treating these AEs as used in the economic model. The manufacturer assumes that Grade 3 and Grade 4 AEs incur the same costs and states that the average cost of treating treatment related AEs in the economic model is derived from costing the AEs reported in PRIMA[32] and EORTC 20981[29, 30] trials. The manufacturer does not state the source/year of the cost data employed.

5.3.8 Perspective, time horizon and discounting

The economic evaluation was undertaken from the perspective of the NHS and PSS in England and Wales. The time horizon set was 25 years. Both costs and benefits were discounted at 3.5% per annum.

5.3.9 Model validation

In response to a clarification request from the ERG, the manufacturer provided details of the economic model validation exercise undertaken. The economic model was checked for internal (technical) validity, convergent validity/corroboration and external and predictive validity.

5.3.10 Results included in manufacturer's submission

Base-case results for the incremental cost per QALY gained are presented in the MS for patients receiving RTX for the first-line maintenance treatment of patients with fNHL compared with observation therapy (Table 21). Total and incremental life years gained (LYG) were also presented in the MS and the incremental cost per LYG is shown in Table 22. The manufacturer demonstrates that RTX as first-line maintenance is cost effective compared with observation in this group of patients.

Table 21 Manufacturer's base-case results: Cost per QALY

Technology	Total costs (£)	Total QALYs	Inc. costs (£)	Inc. QALYs	ICER (£) versus baseline (QALYs)
Observation in 1LM	£66,721	7.207			Baseline
Rituximab in 1LM	£85,403	8.376	£18,681	1.169	**£15,978**
ICER=incremental cost-effectiveness ratio; LYG=life years gained; QALYs=quality adjusted life years; Inc.=incremental; 1LM=first-line maintenance					

Table 22 Manufacturer's base-case results: Cost per LYG

Technology	Total costs (£)	Total LYG	Inc. costs (£)	Inc. LYGs	ICER (£) versus baseline (LYG)
Observation in 1LM	£66,721	9.017			Baseline
Rituximab in 1LM	£85,403	10.288	£18,681	1.271	**£14,697**
ICER=incremental cost-effectiveness ratio; LYG=life years gained; QALYs=quality adjusted life years; Inc.=incremental; 1LM=first-line maintenance					

The MS (MS, Tables 119-121) also presents a series of tables showing very detailed disaggregated costs and benefits. These tables are labelled as: model outputs by clinical outcomes (RTX); model outcomes by clinical outcomes (observation); QALY gain by health state; and summary of cost by health state.

5.3.11 Sensitivity analyses

Scenario analysis and deterministic analyses

The manufacturer investigated the following scenarios:

1. Hazard rate of progression for first-line maintenance is sustained for the life time horizon of the model (best case)

2. Treatment effect of RTX is only maintained for the duration of the PRIMA[32] trial maximum follow up period of 48 months (worst case)

3. Various alternative parametric survival functions to the use of Gompertz in the economic evaluation (Exponential, Log Logistic, Log Normal, Gamma and Weibull) were considered.

The manufacturer also investigated the effect of varying the following parameters: AE costs (+/- 50%); monthly supportive care costs (+/- 50%); RTX administration costs (upper=£267, lower=£176); time horizon (20 years, 30 years); extreme scenario (patients that progress in PF1 transition to death state).

The results of the scenario and deterministic sensitivity analyses are shown in Figure 3. The tornado diagram shows that, in all cases, the ICER remains below £30,000 QALY gained. Only one ICER marginally exceeds £20,000 per QALY gained; this wider variation in the ICER is observed when assumptions regarding the duration of treatment effect are limited to 48 months. The results of the extreme scenario (patients that progress in PF1 transition to death) are presented separately and the estimated ICER is £13,901 per QALY gained.

The ERG considers that the results of the sensitivity analyses undertaken are as anticipated i.e. when expected clinical gains are increased, the size of the ICER falls and vice versa. However, the assumptions inherent in the extreme scenario (patients who progress in PFS1 transition to death) mean that patients die very soon after leaving PFS1, and, as most of the benefits accrue during PFS1, the ERG confirms that this makes very little difference to the size of the ICER.

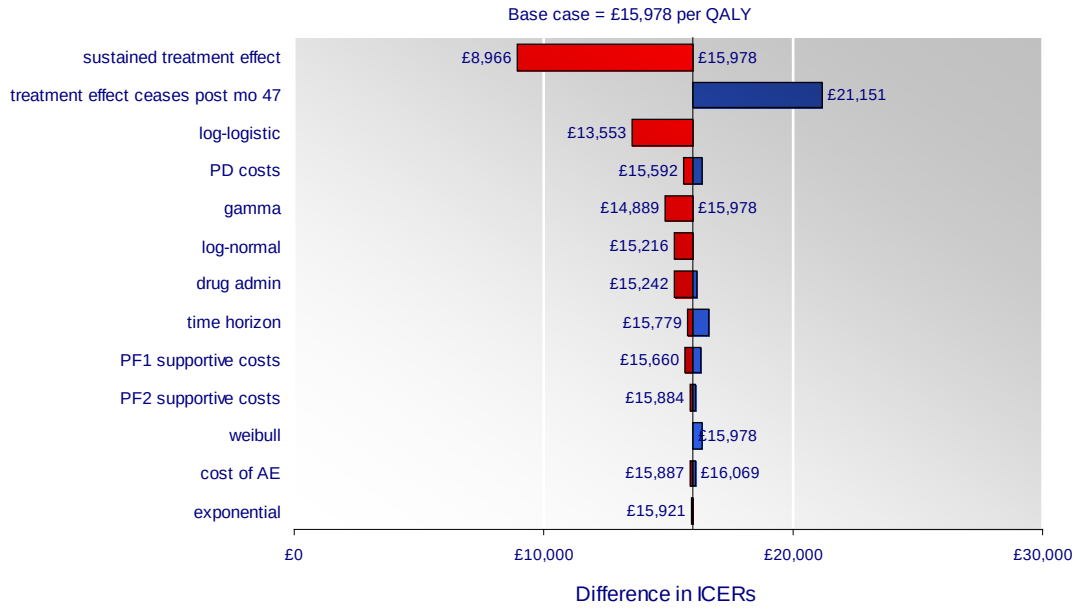

Figure 3 Tornado diagram showing the sensitivity of the model to changes in parameters

Probabilistic sensitivity analysis

Probabilistic sensitivity analysis (PSA) was performed by the manufacturer to assess the robustness of the cost-effectiveness results. All variables were included in the PSA; the only parameters not included in the PSA were age, weight, and height. A Monte Carlo simulation with 2,000 samples was used. The mean resulting ICER was £15,770 per QALY gained. The cost-effectiveness acceptability curve (CEAC) in Figure 4 shows that first-line maintenance therapy with RTX compared with

observation in responding patients with fNHL is highly likely to be cost effective at a threshold of £30,000 per QALY gained.

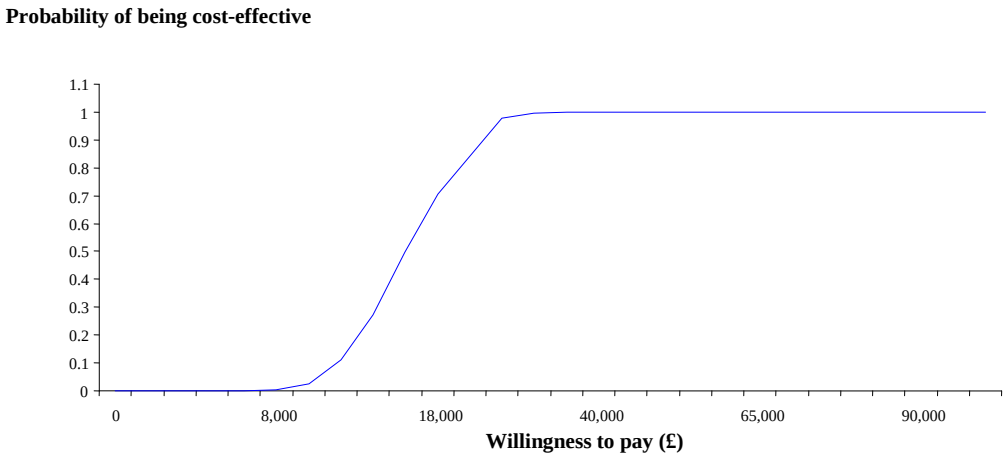

Figure 4 Cost effectiveness acceptability curve (RTX vs observation)

In summary, the manufacturer is confident that the results of the PSA illustrate the robustness of the cost effectiveness of RTX maintenance compared to observation in patients with fNHL responding to first-line induction therapy.

5.3.12 Additional sensitivity analysis undertaken by the manufacturer

In response to a late clarification request from the ERG, the manufacturer conducted a sensitivity analysis in order to demonstrate the robustness of the cost-effectiveness results and to assess the impact of utilising the EFS HR ([redacted] according to the January 2010 cut-off) in the analysis (Table 23); this was found to have a marginal impact on the cost effectiveness of RTX as a first-line maintenance therapy compared with observation.

Table 23 Cost effectiveness analysis of utilising EFS from January 2010

	RTX	Observation
Total cost	£85,574	£66,721
Total QALYs	8.252	7.207
Incremental costs	£18,853	
Incremental QALYs	1.045	
ICER	£18,033 per QALY gained	

5.4 Assessment of the manufacturer's economic model

Table 24 shows how closely the manufacturer's submitted economic evaluation accords with the requirements for a base-case analysis set out in the NICE reference case checklist. There are some key differences between the manufacturer's approach and the NICE reference case checklist. Firstly, the decision problem addressed by the manufacturer is not the same as the question outlined in the scope; the ERG is of the opinion that the manufacturer makes a convincing case for not including zevalin as a comparator. Secondly, the manufacturer's approach to demonstrating clinical effectiveness is considered by the ERG to be flawed; the ERG is of the opinion that the PRIMA[32] data are too immature to be useful in the economic evaluation and that use of clinical data from the EORTC 20981[29, 30] trial is inappropriate as EORTC patients and PRIMA[32] patients are too different in terms of disease status and previous treatments.

Table 25 summarises the ERG's appraisal of the economic evaluation conducted by the manufacturer using the Drummond 10-point checklist. The ERG's main criticism of the submitted economic evaluation is that the key clinical data (from PRIMA[32]) used to populate the economic model are immature and therefore it is not possible to say whether first-line maintenance treatment with RTX compared with observation is or is not cost effective in this group of patients. The ERG considers that although inconsistencies and errors in the submitted model have been identified, their revision would not impact significantly on the size of the manufacturer's base case ICER.

Table 24 NICE reference case checklist

Attribute	Reference case	Does the *de novo* economic evaluation match the reference case?
Decision problem	The scope developed by the Institute	Yes – although the manufacturer makes a valid case for not using one of NICE's recommended comparators
Comparator(s)	Alternative therapies routinely used in the NHS	Yes – as the timing/nature of subsequent therapies is not clear, the ERG cannot comment on whether or not they are routinely used in the NHS
Perspective costs	NHS and Personal Social Services	Yes
Perspective benefits	All health effects on individuals	Yes
Form of economic evaluation	Cost-effectiveness analysis	Yes
Time horizon	Sufficient to capture differences in costs and outcomes	Yes
Synthesis of evidence on outcomes	Systematic review	Limited; the MS uses data from two trials (PRIMA and EORTC). The two trials comprise different patient populations and the ERG does not consider that data from EORTC are meaningful in the economic model. The ERG believes that the data from PRIMA are immature and are not yet sufficiently useful to inform an economic evaluation for this patient group
Outcome measure	Quality adjusted life years	Yes
Health states for QALY	Described using a standardised and validated instrument	Yes
Benefit valuation	Time-trade off or standard gamble	Yes
Source of preference data for valuation of changes in HRQoL	Representative sample of the public	Yes
Discount rate	An annual rate of 3.5% on both costs and health effects	Yes. However, an error in the method used to discount costs and benefits was identified by the ERG – the MS discounted monthly instead of yearly which lead to errors in estimates of costs and benefits
Equity	An additional QALY has the same weight regardless of the other characteristics of the individuals receiving the health benefit	Yes
Sensitivity analysis	Probabilistic sensitivity analysis	Yes

PSS= Personal Social Services; MS= manufacturer submission; RCT= randomised controlled trial; QALYs= quality adjusted life years; PSA= probabilistic sensitivity analysis; ERG= Evidence Review Group; HRQoL= health related quality of life

Table 25 Critical appraisal checklist

Item	Critical appraisal	ERG comment
Was a well-defined question posed in answerable form?	Partial	The question was well-defined but cannot be answered with confidence given the limited clinical data available
Was a comprehensive description of the competing alternatives given?	No	fNHL is a complex disease area and it is difficult to capture all of the lines of therapy over the 25 year period. However, The ERG believes that there is not enough detail describing the nature/timing of treatments in the MS
Was the effectiveness of the programme or services established?	No	Data from PRIMA (key trial) are immature and data from EORTC are inappropriate for decision making
Were all the important and relevant costs and consequences for each alternative identified?	No	As stated above, the timing/nature of subsequent lines of therapy is not clear
Were costs and consequences measured accurately in appropriate physical units?	Yes	Use of NHS Ref Costs 08-09 is appropriate The text of the MS is inconsistent with table text in several places which makes it difficult to be sure of accuracy
Were the cost and consequences valued credibly?	Not always	No, several errors were identified e.g. cost and timing of RTX treatment; mis-use of half cycle correction method
Were costs and consequences adjusted for differential timing?	Yes	An error in the method used to discount costs and benefits was identified by the ERG – the MS discounted monthly instead of yearly which lead to errors in estimates of costs and benefits
Was an incremental analysis of costs and consequences of alternatives performed?	Yes	A lot of detail was provided re incremental analysis in the MS
Was allowance made for uncertainty in the estimates of costs and consequences?	Yes	Yes, deterministic and probabilistic sensitivity analyses were undertaken
Did the presentation and discussion of study results include all issues of concern to users?	No	There were clinical assumptions made in the economic evaluation that were not fully discussed/accounted for in the MS e.g. whether or not timing or frequency of RTX treatment affects duration of expected interval between relapses; full discussion of validity of using EORTC data in economic evaluation was missing in the MS

ERG= Evidence Review Group; QALY= quality adjusted life year; SA= sensitivity analysis; PSA= probabilistic sensitivity analysis; ICER= incremental cost-effectiveness ratio; BSA= body surface area

5.5 Detailed ERG critique of manufacturer's economic model

As an experienced ERG we frequently carry out very detailed revisions of manufacturer's models in order to generate revised economic results for the AC. However, on this occasion, as we consider that there are very high levels of uncertainty arising from shortcomings in the available clinical data, we believe that attempting to make minor modifications to the manufacturer's base-case ICERs would be misleading and inappropriate. The submitted model has a number of important problems, but with credible clinical data could have been modified to assist decision making, but even a 'perfect' model could not overcome the evident weakness of the current evidence for the sustained clinical benefit of RTX in this patient group.

The following sections of the report seek to highlight only those issues which have been identified as indicating sources of potential decision uncertainty, or apparent errors in the design, construction or implementation of the submitted model. In the time available it has not been possible to carry out a comprehensive analysis of all aspects of the model or its parameter values, and the absence of specific comment on any aspect of the model should not be read as implicit endorsement of either the methods or parameter values employed by the manufacturer.

5.5.1 Limitations of clinical effectiveness data

Immaturity of PRIMA data

The reliability of different parts of the Kaplan-Meier PFS survival curves in the post-induction phase of the PRIMA[32] trial (Figure 5) can be gauged by consideration of the proportion of patients censored with respect to PFS at each time point (Figure 6). Analyses indicate that for the first 2 years this indicator is very low (<3%) and is consistent with the absence of right-censoring (patient data curtailed by data cut off, rather than progression, death or drop-out of the trial) in this period. However, beyond 800 days censoring increases very rapidly to reach more than 70% (RTX) and 50% (observation) by 1600 days. Such high levels of right-censoring provide considerable opportunities for unpredictable bias to occur, especially as the number of patients remaining at risk falls to low levels. The implication of these findings is that the reliability of the Kaplan-Meier survival estimate up to 800 days is very good, but for later times it is much less trustworthy and this has important implications for the use of long-term projective modelling to populate the decision model.

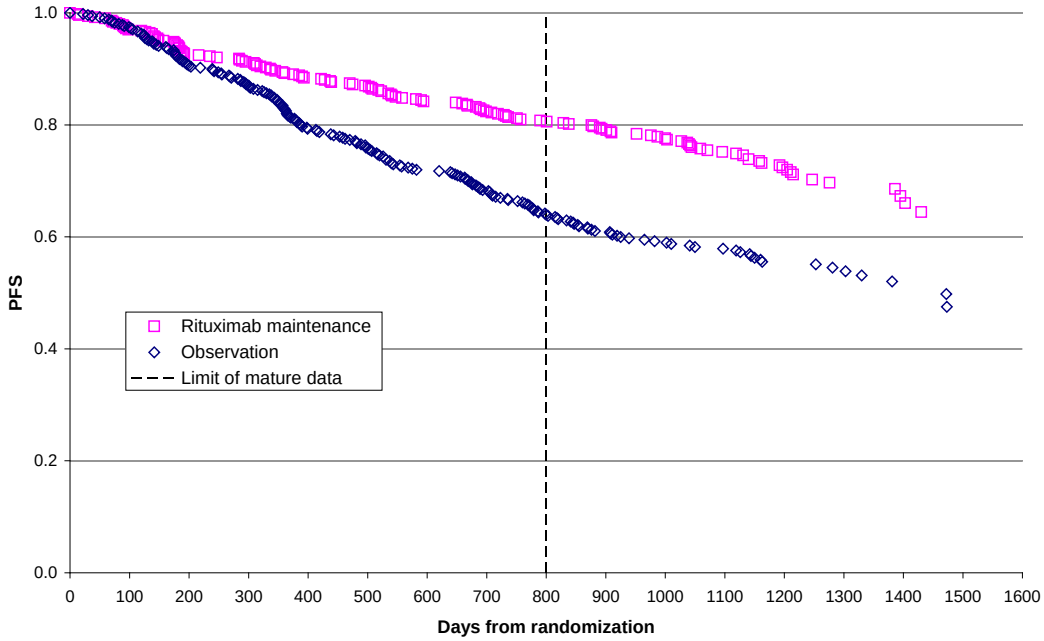

Figure 5 Kaplan-Meier analysis of PFS following second randomisation in PRIMA

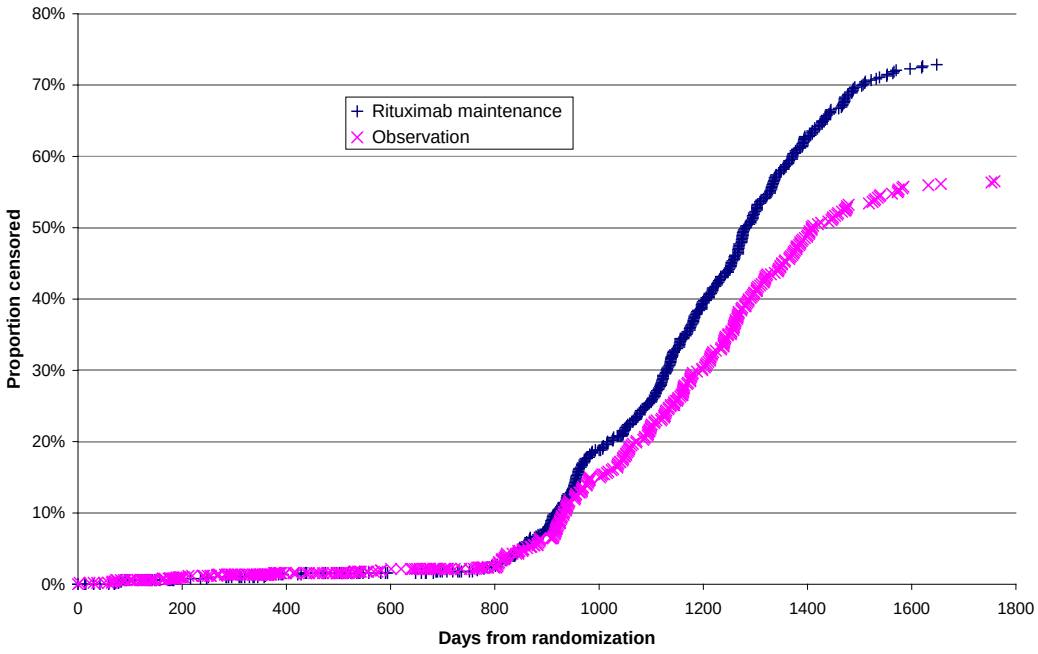

Figure 6 Cumulative proportion of PRIMA patients censored for PFS by time from second randomisation

The ERG has used data prior to 800 days to illustrate the difficulty of determining an appropriate choice of parametric function to project PFS throughout the duration of the model (25 years). Each of the six functions tested by the model authors has been fitted to these data and the results are shown graphically in Figure 7 (observation arm) and Figure 8 (RTX maintenance arm). The use of statistical measures of correspondence (AIC or BIC) are of little value in distinguishing between these options, since all options match the data reasonably closely in the first 2 years but diverge markedly thereafter. The wide range of these long-term projections suggests that the choice must be made on other than statistical grounds. Most important is inherent clinical plausibility – in the light of our current experience of the normal course of the disease in these patients, what level (if at all) can it be judged will be the proportion of patients in the RTX maintenance arm who will be alive and progression/relapse free after 25 years: 10% (exponential model), 47% (Gompertz model) or none of these? Our clinical advisor confirms that a very small proportion of patients is expected to be alive and progression/relapse free after 25 years.

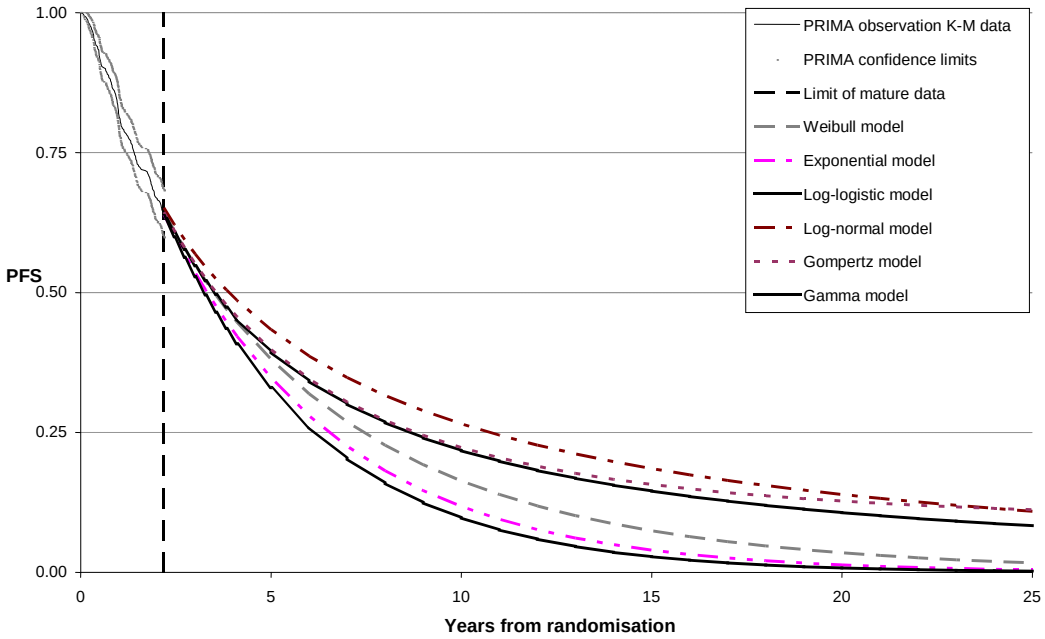

Figure 7 Six parametric survival functions fitted to PRIMA data up to 800 days from randomisation: observation arm

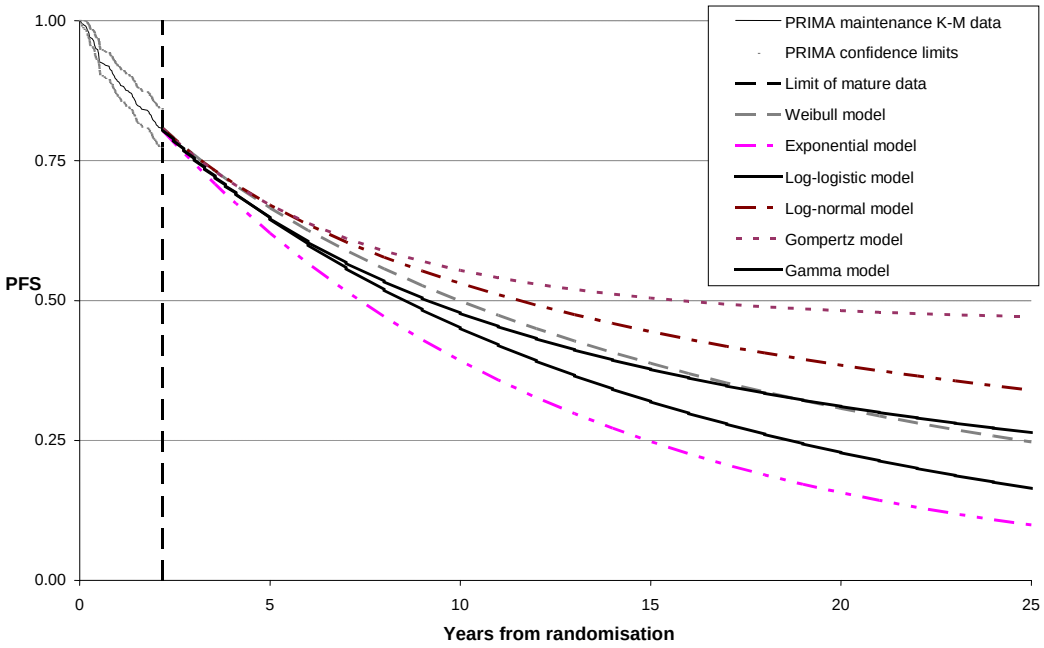

Figure 8 Six parametric survival functions fitted to PRIMA data up to 800 days from randomisation: RTX maintenance arm

This is merely an illustration; in practice the model authors impose an arbitrary time limit on the duration of direct benefit attributable to RTX of 6 years. Nonetheless, it demonstrates how the lack of long-term mature trial data renders any attempts at mathematical projection distinctly unconvincing, so that the gains obtained by the 'best performing' function (Gompertz - as used in the manufacturer's base case) may lead to much higher estimates of health gain than the most conservative (exponential) option without any demonstrable justification.

5.5.2 Model design and structure

Health states and pathways

It appears that the model designers, in attempting to simplify the complexities of frequently complex and extended disease histories, may have omitted important additional states and pathways from the model, potentially leading to distorted cost-effectiveness results. Following progression in the maintenance/observation state, all surviving patients move immediately into a second PFS state having presumably undergone a second successful induction treatment. By omitting this second treatment (and in effect treating it as a zero-duration event occurring at the same time as disease progression), this model structure creates four anomalies:

- the time spent in second-line therapy is missing from calculations of OS, and subsequent costs and benefits may not be correctly discounted;

- the cost of the second-line therapy cannot be correctly located in time, and is all attributed to the first cycle of the second PFS state;

- the disutility of disease progression and undergoing CTX are not accounted for;

- there is no pathway in the model to allow some patients to fail to achieve a remission at this second course of induction therapy.

Following a second-line maintenance/observation PFS period, relapsed patients move to a progressive disease state in which a third-line of anti-lymphoma treatment is embedded. This does not allow any further rounds of treatment to occur despite well attested registry studies[17, 55, 56] showing that more than four phases of active treatment are not uncommon.

These difficulties highlight a general issue relating to conditions with longer/lifetime duration and complex pathways. In these circumstances it becomes increasingly difficult to represent adequately real-life patterns of care within a Markov framework, which requires the definition of essentially homogeneous health states in which patients share common risks, utility and treatment costs regardless of their prior history. There is a strong rationale for employing patient level simulation methods in such situations.

Principal model driver

The duration of PFS after a first successful course of induction therapy with or without maintenance is the principal driver of health gain in the model, and is informed by the primary outcome of the PRIMA[32] trial. However, PFS in clinical trials is frequently determined retrospectively from laboratory and/or radiological evidence rather than clinically. In the PRIMA[32] trial there is some indication that EFS (incorporating the time to next anti-lymphoma treatment) may be somewhat reduced compared to PFS. In view of the primary role of the timing of each phase of induction therapy, the ERG is of the view that the 'time to next anti-lymphoma treatment' is the more appropriate outcome variable to drive the timing of transitions through model pathways, and better reflects clinical reality.

Second source of effectiveness information

The ERG is mindful of the difficulty involved in populating lifetime decision models with credible data. In this case the manufacturer has drawn on an earlier study of RTX maintenance treatment in patients with fNHL (EORTC 20981[29, 30]) to represent the risk of progression during second-line RTX maintenance. This is understandable, but is problematic since the patients in the EORTC[29, 30] trial are at a different stage in their disease career (second or third-line induction therapy, rather than chemo-naïve patients as in the PRIMA[32] trial), and, in addition, had not received RTX previously. The effect of sequential filtering out of patients who fail to respond to various stages of induction treatment is very likely to give rise to important differences in patient characteristics, and in their propensity to benefit from subsequent courses of RTX as a result. Thus the use of these data to calibrate the submitted model must be considered questionable, since it involves strong assumptions of equivalence of effect which cannot be empirically validated from existing evidence.

5.5.3 Implementation of the model

Estimating deaths

The estimation of deaths in the first PFS state is problematic in the submitted model. It relies on an estimated monthly death rate derived from the PRIMA[32] trial, and also the published mortality rates by age and sex for England and Wales. The model logic uses whichever of the two estimates is greater, except where this appears to exceed the estimated total number of PFS events for the period when a questionable 'adjustment' is applied to prevent negative numbers being generated. In principle, the ERG believes that the most appropriate approach to incorporating mortality into the model would be to use the sum of the lymphoma rate and the age-specific

population rate (which is a very close approximation to non-lymphoma mortality). In addition, the mortality rate should normally be applied to the number of survivors at the start of the relevant period (month) and not to a mid-cycle average as in the submitted model. Overall this feature of the model appears to be unsatisfactory.

Eligibility for second-line treatment

The submitted model uses data from the PRIMA[32] trial to estimate the proportion of patients failing from the first PFS period (observation or RTX maintenance) but who did not progress to second-line induction therapy. These proportions were calculated relative to the whole randomised population, but are applied in the model only to those patients still alive at the end of PFS. The ERG's correction of this anomaly has only a minor effect on the base case estimates, increasing both incremental costs and outcomes, and raising the ICER by £19 per QALY gained.

Estimation of event rates from EORTC 20981 results

The manufacturer has estimated monthly HRs to govern PFS and OS in second-line RTX maintenance from the Kaplan-Meier plots for the EORTC 20981[29, 30] trial. The method used assumes that a simple exponential model will suffice in all cases. Ordinary least squares (OLS) regression is used to fit a survival model to digitised data for the whole trial period, and the monthly risk parameter is calculated as the average rate over the first 12 months. However examination by the ERG of the fitted models compared to the trial data reveals two problems with this approach:

- the OLS models allow a variable starting value for survival to be computed, despite the constraint on Kaplan-Meier data to commence with 100% of patients event-free at baseline;

- the cumulative hazard plots for several of the data sets (especially for PFS) show strong evidence of non-linearity, with a period of early high risk, followed by a lower long-term risk.

The combined effect of these problems is to render several of the fitted models clearly inaccurate and inappropriate for the purpose of calibrating the manufacturer's model for the second-line maintenance phase.

Discounting

The manufacturer has failed to implement discounting correctly according to UK practice (i.e. applied annually). The ERG incorporated a change from monthly to annual discounting. This change increased the incremental overall discounted cost

per patient by £736 (+3.9%) and incremental discounted QALYs per patient by 0.019 (+1.6%), resulting in the ICER increasing by £370 per QALY gained.

Mid-cycle correction

A feature of health state models using a fixed cycle length is the need to convert point estimates (e.g. of survival) into period estimates (i.e. the average exposure to risk of an event during a period). This requires the use of a mid-cycle correction to be applied at appropriate points in the model. However, it is essential to avoid the possibility of confusion over which variables require correction and at which points in the logic of the model to apply the correction. The ERG has detected at least one instance in the submitted model of a problem in implementing a mid-cycle correction; a variable (deaths in PFS) has been created by averaging a second variable which itself has already had a mid-cycle correction applied to it i.e. a case of double adjustment. However, the estimated number of deaths should not be subject to mid-cycle correction at all. The extent of this error is difficult to assess without carefully restructuring the model to ensure the separation of primary time point estimates from secondary period estimates of costs and outcomes.

Adverse events

The costs of treating AEs have been estimated for Grade 3 and 4 events recorded in the PRIMA[32] trial, for experience in both first-line maintenance/observation and also in second-line maintenance/observation. All costs are aggregated into a variable which is applied in the first cycle of the model, and is therefore undiscounted. Since the PRIMA[32] trial data are very immature it is likely that the second-line AE costs will be substantially underestimated in the model since the majority of patients had not yet progressed from the first-line maintenance/observation state at the time of data cut off.

No estimate has been included in the model for the disutility associated with Grade 3 and 4 AEs, which potentially favours RTX, as RTX patients have higher AE incidence rates.

Utility values

In the MS the manufacturer uses the following utility values from the Oxford Outcomes [49] study for PFS: PFS1= 0.88 (disease free); PFS2 = 0.79 (remission/full response to therapy). The ERG considers that there should be no difference in the utility values used in the model to describe PFS1 and PFS2 as both groups of patients are in "remission/full response". When corrected by the ERG, this revision has the

effect of reducing the QALY gain in PRIMA[32] PFS by more than 10% and will therefore increase the ICER by approximately 11%.

Cost of rituximab maintenance treatment

The manufacturer estimates RTX costs using an overall average BSA figure (mean BSA 1.84m^2) to estimate the cost per dose of RTX, without adjusting for the wide range of BSA values in the population and gender-specific BSA differences (mean BSA: females 1.71 m^2, males 1.95 m^2). When recalculated by the ERG, taking these factors into account, the mean cost per dose increases slightly from £1,222.39 to £1,281.52 (+4.8%). This increases the incremental total discounted cost per patient by £1,032 (+5.5%), and the ICER by £882 per QALY gained. The ERG found little evidence in the PRIMA trial results supporting substantial non-compliance with RTX therapy, and so conclude that no adjustments are appropriate for missed doses.

Timing of rituximab maintenance doses

Rituximab first-line maintenance treatment is administered 12 times at 8 week intervals. This means that the last dose occurs 88 weeks (20.2 months) after the first dose. In the submitted model the cost of the 12 doses is spread evenly over 24 months which is equivalent to assuming half a dose mid-way through each month. In fact the dosing schedule leads to an uneven dosing across the monthly model periods, with seven doses in the first year and five doses in the second year (when discounting applies). Correction of the model by the ERG affects both the discounted cost of RTX and the discounted cost of administration. The incremental overall discounted cost per patient is thereby reduced by £566 (3.0%), and the ICER is reduced by £484 per QALY gained.

Health state costs

The description provided by the manufacturer of the derivation of supportive costs for patients in PFS or with progressive disease is not always clear, and some of the assumptions underlying the estimates are questionable. Nonetheless, since the same parameter values are used in both arms of the model, and the monthly costs are generally quite low, the impact of inaccuracy in these parameter values is limited, as evidenced by sensitivity analyses.

5.5.4 Model sensitivity

Two aspects of the manufacturer's model, which appear to be influential in governing the cost-effectiveness results, are worthy of note. Both affect the

estimated health gain arising from RTX maintenance treatment after first successful induction treatment: the age of patients, and the assumed duration of clinical benefit.

Patient age

As confirmed by our clinical advisor, the age at which a patient is diagnosed with fNHL and begins their first course of CTX is of great importance in determining the potential benefit which may accrue from use of a novel regimen. Those beginning treatment at an advanced age are subject to increasingly elevated all-cause mortality risks which inevitably limit the scope for gains in both OS and PFS. Therefore in such long-term incurable conditions, age is always a potential factor predisposing to poorer cost effectiveness for any treatment. There may also be additional problems in that general debilitation could inhibit response to treatment or shorten its duration.

To test this latter hypothesis in the case of RTX maintenance treatment, the ERG requested additional results via the clarification process from the PRIMA[32] trial to allow a comparison of clinical effectiveness between three age-based subgroups. The manufacturer provided these data in the form of numbers of PFS events and estimated odds ratios for RTX vs observation for patients aged under 44 years, 44-64 years and 65+ years.

Despite the immaturity of the PRIMA[32] data, and the unsophisticated nature of the analysis, there appears to be evidence of an emerging trend indicating a reduction of clinical effect as patient age increases – a curvilinear trend in OR, equivalent to a linear trend in relative risk. To illustrate the sensitivity of model results to this effect, the HR of PFS in the base case model was adjusted by the ERG to reflect specific patient ages, indicating that the combined effect of increasing mortality and reducing effectiveness can alter the estimated ICER substantially (Table 26).

Table 26 Sensitivity of manufacturer's base case cost-effectiveness estimates to patient age

Patient age at induction (years)	Total discounted cost per patient		Total discounted QALYs per patient		Incremental		Estimated ICER (£/QALY)
	RTX	Obs	RTX	Obs	Cost per patient	QALYs per patient	
30	£86,090	£67,787	9.187	7.317	£18,304	1.870	£9,790
40	£86,364	£67,759	8.902	7.314	£18,605	1.588	£11,716
50	£86,266	£67,404	8.605	7.278	£18,862	1.327	£14,214
60	£84,280	£65,840	8.137	7.116	£18,440	1.021	£18,055
65	£81,815	£63,848	7.761	6.910	£17,967	0.852	£21,099
70	£77,875	£60,489	7.241	6.563	£17,386	0.677	£25,668
75	£72,508	£55,695	6.578	6.068	£16,806	0.511	£32,913
80	£66,618	£50,085	5.868	5.486	£16,533	0.382	£43,306

Assumed duration of clinical benefit

As a consequence of the immaturity of the PRIMA[32] trial, it proved necessary for the manufacturer's model to be designed to include a parameter governing the maximum period over which RTX maintenance therapy could be expected to provide direct benefit (i.e. reduced risk of disease progression). Since RTX was not given to patients beyond 2 years, this represents a minimum value for this parameter. In the base case, the manufacturer sets this period to 6 years but without referring to any supporting evidence. Two alternatives are offered: 4 years (equivalent to the maximum time over which any patients had been observed within the PRIMA[32] trial), and 40 years (equivalent to a lifetime).

In Table 27, the base case model is used to illustrate the effect of different values of this effectiveness limit on the model ICER, indicating that if a strict interpretation of data maturity (800 days as discussed previously) were applied, it would be sufficient to lead to an ICER greater than £30,000 per QALY gained.

Table 27 Sensitivity of manufacturer's base case model to variations in the assumed maximum period of rituximab effectiveness

Max period RTX is effective (months)	Total discounted cost per patient		Total discounted QALYs per patient		Incremental		Estimated ICER (£/QALY)	Scenario rationale
	RTX	Obs	RTX	Obs	Cost per patient	QALYs per patient		
26	£86,865	£66,721	7.832	7.207	£20,143	0.625	£32,230	Mature data (800 days)
36	£86,538	£66,721	7.967	7.207	£19,816	0.760	£26,079	
48	£86,130	£66,721	8.124	7.207	£19,409	0.918	£21,151	Limit of trial data
60	£85,753	£66,721	8.260	7.207	£19,032	1.053	£18,071	
72	£85,403	£66,721	8.376	7.207	£18,681	1.169	£15,978	Base case model
480	£82,218	£66,721	8.935	7.207	£15,496	1.728	£8,966	Lifetime

However, it is important to note the nature of this restriction as it has been implemented within the manufacturer's model; at the end of the assumed 'effective period' the hazard governing the PFS projective model reverts to the hazard of the comparator arm, but from a higher absolute survival level. This approach ensures that the two survival curves will never converge within a finite time, ensuring that the RTX arm continues to accumulate survival gains long after the supposed limit of clinical effectiveness. This assumption therefore precludes an alternative process observed in some clinical trials, whereby clinical gains begin to decay following the end of active treatment, until the two survival curves converge to the same long-term trajectory after a few months or years. In simple terms the difference can be characterised as follows:

- the manufacturer's approach assumes that the clock governing the disease process is 'turned back' for patients on RTX by several years, so that it never catches up with the untreated patients;

- the alternative process assumes that the disease clock is slowed down by RTX for several years, but then accelerates when the effect wears off, and eventually catches up with the disease in the untreated patients.

The difference in health gain between these two modelling assumptions is difficult to estimate from the current model. However, in the manufacturer's base-case more than 72% of the estimated PFS gain arises after the 4 year point; there is therefore considerable scope for reduction in incremental outcomes if the initial PFS advantage

is lost progressively Thus the estimated ICER could increase by as much as three times, depending on the time over which the gap between the PFS curves disappears.

5.5.5 Usefulness of model results

None of the detailed implementation issues described above (section 5.5.3) appears likely to have a major impact on the estimated ICER for use of RTX as first-line maintenance therapy, either individually or in combination. Several of the more structural problems (section 5.5.2) could not be corrected by the ERG, but it is unlikely that they will impact on any decision based on an assessment of cost effectiveness. However, important sources of substantial uncertainty remain which impact detrimentally on the apparent cost effectiveness of RTX maintenance therapy.

The main source of uncertainty relates to the immature nature of the primary data source – the PRIMA[32] trial. The submitted model projects future benefits in terms of increased patient time in PFS, and this is the dominant driver of cost effectiveness. In the manufacturer's base case results the model estimates the mean (undiscounted) survival for patients without maintenance therapy as 11.44 years, and 13.38 years for those with RTX maintenance: a gain of 1.94 years. The model also estimates the mean time spent in PFS as 8.64 years (observation only) and 10.65 years (RTX maintenance): a gain of 2.01 years. Virtually all this benefit is generated in the first PFS phase of the model. This implies that PFS gains achieved directly by extending the first-line induction response are translated almost entirely (96.6%) into OS gains. This must be considered a 'best possible' scenario which requires strong supportive evidence from clinical trials before it can be accepted.

At present there is no unequivocal evidence from any clinical trial or meta-analysis of RTX maintenance treatment of patients with fNHL for any significant OS gains, despite strong evidence of PFS gains. The immaturity of the PRIMA[32] trial data compounds this problem, since the extent of PFS gain cannot be estimated directly, but only by projective modelling. It should be noted that the PFS advantage from first-line RTX maintenance that is measurable directly from the mature trial data (up to 800 days from randomisation using the latest clinical data available) amounts to no more than 60 days.

The ERG also anticipates that the size of the ICER will be affected by the age of the patients in the economic model and the assumed duration of clinical benefit from RTX. Depending on the assumptions made, the manufacturer's base case ICER could increase substantially.

Given the uncertainties discussed, the question must be explored of whether *any* reliance should be placed on results obtained from the submitted model in seeking to resolve the decision problem.

There are, however, some elements of the problem in which we can have confidence:

- the additional cost to the NHS of RTX first-line maintenance therapy is not in doubt (about £17,700 per patient);

- in the early 'mature' part of the trial data set there is strong reliable evidence of reduced risk of progression/relapse at least up to about 800 days;

- subsequent phases of treatment are unlikely to have more than a minor impact on the overall incremental cost per patient.

If the manufacturer's estimate of PFS gains (using Gompertz projection with benefit consolidated but not extended beyond 6 years) were accepted for the purpose of illustration, we may consider the likely impact of different rates of conversion of PFS gain into OS gain on the estimated ICER. The results shown in Table 28 indicate that at least 50% of PFS gain would need to result in OS gain to achieve an ICER below £30,000 per QALY gained. If a less generous PFS projection method (i.e. not Gompertz) were used, the required conversion rate would probably be higher.

Table 28 Illustrative estimated ICERs based on manufacturer's base case, adjusted for a range of possible conversion factors for PFS gain into OS gain

Conversion rate PFS gain to OS gain	Approximate ICER (£/QALY) estimated by the ERG
96.6%	£15,978
90%	£17,155
80%	£19,300
70%	£22,057
60%	£25,733
50%	£30,880
40%	£38,600
30%	£51,466
20%	£77,199
10%	£154,399

ERG conclusions on submitted cost-effectiveness evidence

The ERG considers that direct use of the model results is too heavily affected by extensive and often unquantifiable uncertainty around the central claims of clinical benefit to be useful for decision making i.e. the ERG concludes that the clinical data

available make it impossible to compare the cost effectiveness of first-line RTX maintenance vs observation in patients with fNHL with any confidence.

5.6 Summary of cost-effectiveness evidence

5.6.1 Economic evaluation results

Base-case: Manufacturer

- The manufacturer reports an ICER of £15,978 per QALY gained for the comparison of first-line maintenance RTX vs observation in patients with fNHL.
- Results of the PSA conducted by the manufacturer suggest that, based on the assumptions made and evidence available, RTX as a first-line maintenance therapy has a very high probability of being cost-effective at a WTP of £30,000 per QALY gained compared with observation.

Base-case: ERG

- The submitted model projects future benefits in terms of increased patient time in PFS, and this is the dominant driver of cost effectiveness in the economic model. The ERG is of the opinion that the PFS data from the primary data source – the PRIMA[32] trial – are immature. The ERG has therefore chosen *not* to revise the manufacturer's estimate of the base case ICER.

5.6.2 Economic issues and uncertainties

- The ERG has identified both structural (omitted health states and pathways, use of PFS instead of EFS as key outcome, use of non-comparable clinical effectiveness evidence from second clinical data source) and implementation anomalies (e.g. non-standard use of discounting method, inaccurate costing of RTX treatment, underestimation of AEs) in the economic model. However, the ERG highlights that none of the issues described appear likely to have a serious impact on the estimated ICER for use of RTX as first-line maintenance therapy, either individually or in combination.
- The ERG considers that lack of mature clinical effectiveness data from the PRIMA[32] trial combined with the sensitivity of the base case ICER to several important model assumptions (e.g. age of the patients, assumed duration of benefit from RTX) indicate much greater uncertainty in model cost-effectiveness results than is suggested by the manufacturer's sensitivity analyses.
- The ERG concludes that the lack of mature clinical data makes it impossible to compare the cost effectiveness of first-line RTX maintenance vs observation in fNHL patients with any confidence. Whether or not first-line RTX maintenance vs observation is or is not cost effective compared to observation cannot be determined from the evidence submitted.

6 DISCUSSION

The manufacturer presents a case for the use of RTX as a maintenance treatment for patients with high tumour burden fNHL whose disease has responded to first-line induction therapy with CTX.

The systematic review carried out by the manufacturer identified a single relevant RCT (PRIMA[32]) which compared the use of RTX monotherapy with observation. The PRIMA[32] trial was an open label, international, multi-centre RCT. The results of PRIMA[32] are as yet unpublished.

Whilst the results presented in the MS appear to be promising, the early closure of the PRIMA[32] trial, combined with limited follow-up, has resulted in an immature dataset; too few patients have experienced events to draw any firm conclusions on clinical and cost effectiveness from the currently available data.

The ERG notes that a number of authors have commented on the problems associated with data from trials that are terminated early.[57-60] A recent meta-analysis[57] compared the reported results of 91 RCTs that were halted early for benefit with 424 similar RCTs that ran to full term. The authors found large differences in treatment effect size between trials that were stopped early and similar trials that ran their full course. This was true regardless of the methodological quality of the trial or the presence of statistical stopping rules. One implication of this finding is that early closure of trials can lead to exaggerated treatment effects that would not be borne out in the longer term. Other important implications pertinent to the early termination of cancer trials are summarised by Trotta and colleagues[58] in their 2008 analysis of trials that were stopped early for benefit:

> *"if a trial is evaluating the long-term efficacy of a treatment of conditions such as cancer, short-term benefits, no matter how significant statistically, may not justify early stopping. Data on disease recurrence and progression, drug resistance, metastasis, or adverse events, all factors that weight heavily in the benefit/risk balance, could easily be missed. An early stop may reduce the likelihood of detecting a difference in overall survival (the only relevant end point in this setting) because of the small sample, the possibility of crossing-over the experimental drugs, and contamination with other treatments".*

The manufacturer has provided evidence for the efficacy and safety of re-treatment with RTX. However, the ERG is unable to comment on the potential impact that RTX, when used as a first-line maintenance treatment, will have on the efficacy and safety of subsequent treatments with RTX for patients with fNHL.

The ERG has scrutinised the cost-effectiveness evidence submitted by the manufacturer in support of RTX as a first-line maintenance therapy for patients with fNHL. Several implementation and structural concerns regarding the submitted economic model have been identified by the ERG. However, the ERG report acknowledges that if the suggested ERG corrections and/or modifications are made to the submitted economic model by the manufacturer, the size of the ICER is unlikely to be substantially affected.

The ERG has raised other concerns related to the economic models which are not so easily dealt with. Firstly, the ERG is of the opinion that the clinical data currently available from the PRIMA[32] trial are too immature to be considered useful for decision making regarding the use of RTX in this group of patients. The PRIMA[32] trial was stopped before median PFS data were available and, at the time of writing, median PFS for the RTX arm of the trial is still not estimable; the ERG highlights that it is the PFS data which drive the results of the economic model. In the submitted economic model, the manufacturer chose to use the Gompertz survival function to model PFS and implicitly employed a conversion rate of 96.6% from PFS gain to OS gain; in doing so, the benefits of RTX as a first-line maintenance therapy are maximised. Whether or not the estimated survival gains (and therefore estimated ICERs) will be confirmed by future data from PRIMA[32] and/or other studies is unknown. As a result, the ERG is of the opinion that it is currently impossible to compare the cost effectiveness of first-line RTX maintenance vs observation in fNHL patients with any confidence.

Secondly, the manufacturer's estimate of the base case ICER is low (<£16,000 per QALY gained); however, the ERG has shown that the economic model submitted is very sensitive to two assumptions (age of the patients and assumed duration of clinical benefit). This means that, depending on the assumptions governing the economic model, significantly increased ICERs are a genuine possibility. For example, the ERG considers that, by simply reducing the assumed duration of clinical benefit from RTX, the manufacturer's base case ICER may well increase substantially.

In conclusion, the ERG is of the opinion that the cost effectiveness of RTX as a first-line maintenance therapy vs observation cannot be determined from the evidence submitted by the manufacturer. The ERG eagerly awaits the publication of more mature data from the PRIMA[32] trial as without this (or equivalent long-term data

from other compatible trials) it may be impossible to address the question of whether RTX is a cost-effective treatment in this group of patients.

6.1 Implications for research

In this submission, the dataset presented was considered by the ERG to be too immature for a reliable assessment of clinical efficacy or cost effectiveness. Whilst the data presented appear to be promising, a longer follow-up period in which more events have accrued is required to inform decision-making.

Due to the paucity of long-term evidence for the continued benefit of RTX and its safety, monitoring and surveillance of patients who have been treated and re-treated with RTX are therefore necessary.

7 REFERENCES

1. Cancer Research. Non-Hodgkin lymphoma (NHL) statistics - UK2010: Available from: http://info.cancerresearchuk.org/cancerstats/types/nhl/index.htm?script=true.
2. The Non-Hodgkin's Lymphoma Classification Project. A clinical evaluation of the International Lymphoma Study Group classification of non-Hodgkin's lymphoma. Blood. 1997;**89**:3909-18.
3. Shipp MA, Mauch PM, Harris NL, editors. Non-Hodgkin's lymphomas. 5th ed. Philadelphia: Lippincott-Raven; 1997
4. Horning SJ, Rosenberg S. Natural history of initially untreated low-grade non-Hodgkin's lymphoma. New Engl J Med. 1984;**311**:1471-5.
5. Lister T. The management of follicular lymphoma. Ann Oncol. 1991;**2**(Supp 2):131-5.
6. Swenson WT, Wooldridge JE, Lynch CF, Forman-Hoffman VL, Chrischilles E, Link BK. Improved survival of follicular lymphoma patients in the United States. Journal of Clinical Oncology. 2005;**23**:5019-26.
7. Liu Q, Fayad L, Hagemeister FB, Alma Rodriguez M, Younes A, Pro B, et al. Stage IV indolent lymphoma: 25 years of treatment progress. Blood. 2003;**102** (398a abstr 1446).
8. Dillman RO, Chico S. Improved survival of lymphoma patients after introduction of rituximab. Blood. 2005;**106**(abstr 4650).
9. Fischer RI, LeBlanc M, Press OW, Maloney DG, Unger JM, Miller TP. New treatment options have changed the survival of patients with follicular lymphoma. Journal of Clinical Oncology. 2005;**23**(33):8447-52.
10. Gall EA, Mallory TB. Malignant lymphoma: A clinico-pathologic survey of 618 cases. Am J Pathol. 1942;**18**:281-429.
11. Carbone PP, Kaplan HS, Musshoff K, Smithers DW, Tubiana M. Report on the Committee on Hodgkin's disease staging classification. Cancer Res. 1971;**31**(11):1860-1.
12. The International non-Hodgkin's Lymphoma Prognosis Factors Project. A predictive model for aggressive non-Hodgkin's lymphoma. New Engl J Med. 1993;**329**:987-94.
13. Morschhauser F, Radford J, Van Hoof A, Vitolo U, Soubeyran P, Tilly H, et al. Phase III trial of consolidation therapy with yttrium-90-ibritumomab tiuxetan compared with no additional therapy after first remission in advanced follicular lymphoma. Journal of Clinical Oncology. 2008 Nov 10;**26**(32):5156-64.
14. National Institute for Health and Clinical Excellence. Rituximab for the first line maintenance treatment of follicular non-Hodgkin's lymphoma: final scope. London: NICE; 2010 [accessed 2010 August]; Available from: http://guidance.nice.org.uk/TA/Wave19/59/Scope/pdf/English.
15. Dreyling M. Newly diagnosed and relapsed follicular lymphoma: ESMO clinical recommendations for diagnosis, treatment and follow up. Ann Oncol. 2010; **21**(supp 5):v181-3.
16. McManus M, Hoppe R. Is radiotherapy curative for stage I and II low grade follicular lymphoma? Journal of Clinical Oncology. 1996;**14**(4):1282-90.
17. Gallagher CJ, Gregory WM, Jones AE, Stansfeld AG, Richards MA, Dhaliwal HS, et al. Follicular lymphoma: prognostic factors for response and survival. Journal of Clinical Oncology. 1986;**4 (Oct)**:1470-80.
18. Wahl R, Kaminski M, Zelenetz A, Vose J, Press O, Goldsmith S, et al. Each subsequent therapy results in diminishing response rate and duration of response in low grade or transformed low grade non-Hodgkin's lymphoma.

Proceedings of the American Society of Clinical Oncology [serial on the Internet]. 2001; 20.
19. Genactis. Patient record survey 2010.
20. Buske C, Hoster E, Dreyling M, Forstpointner R, Kneba M, Schmitz N, et al. Rituximab in combination with CHOP in patients with follicular lymphoma: analysis of treatment outcome of 552 patients treated in a randomized trial of the German Low Grade Lymphoma Study Group (GLSG) after a follow up of 58 months. Blood. 2008;**112**(11):901.
21. Herold M, Haas A, Srock S, Neser S, Al-Ali KH, Neubauer A, et al. Rituximab added to first-line mitoxantrone, chlorambucil, and prednisolone chemotherapy followed by interferon maintenance prolongs survival in patients with advanced follicular lymphoma: An East German Study Group hematology and oncology study. Journal of Clinical Oncology. 2007;**25**(15):1986-92.
22. Hiddemann W, Kneba M, Dreyling M, Schmitz N, Lengfelder E, Schmits R, et al. Frontline therapy with rituximab added to the combination of cyclophosphamide, doxorubicin, vincristine, and prednisone (CHOP) significantly improves the outcome for patients with advanced-stage follicular lymphoma compared with therapy with CHOP alone: results of a prospective randomized study of the German Low-Grade Lymphoma Study Group. Blood. 2005;**106**(12):3725-32.
23. Marcus R, Imrie K, Belch A, Cunningham D, Flores E, Catalano J, et al. CVP chemotherapy plus rituximab compared with CVP as first-line treatment for advanced follicular lymphoma. Blood. 2005;**105**(4):1417-23.
24. Marcus R, Imrie K, Solal-Celigny P, Catalano JV, Dmoszynska A, Raposo JC, et al. Phase III study of R-CVP compared with cyclophosphamide, vincristine, and prednisone alone in patients with previously untreated advanced follicular lymphoma. Journal of Clinical Oncology. 2008;**26**(28):4579-86.
25. Salles G, Mounier N, de Guibert S, Morschhauser F, Doyen C, Rossi JF, et al. Rituximab combined with chemotherapy and interferon in follicular lymphoma patients: results of the GELA-GOELAMS FL2000 study. Blood. 2008;**112**(13):4824-31.
26. Schulz H, Bohlius JF, Trelle S, Skoetz N, Reiser M, Kober T, et al. Immunochemotherapy with rituximab and overall survival in patients with indolent or mantle cell lymphoma: A systematic review and meta-analysis. J Natl Cancer Inst (Bethesda). 2007;**99**(9):706-14.
27. National Institute for Health and Clinical Excellence. Rituximab for the treatment of follicular lymphoma: TA110. London: NICE; 2006 [accessed 2010 25th August]; Available from: http://guidance.nice.org.uk/TA110/Guidance/pdf/English.
28. Scottish Medicine Consortium. rituximab, 100mg and 500mg concentrate for solution for infusion (MabThera®) 2008: Available from: http://www.scottishmedicines.org.uk/files/rituximab%20(MabThera)%20Final%20August%202008.doc%20for%20website.pdf.
29. van Oers M, 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.
30. van Oers MHJ, van Glabbeke M, Giurgea L, Klasa R, Marcus RE, Wolf M, et al. Rituximab maintenance treatment of relapsed/resistant follicular non-Hodgkin's lymphoma: long-term outcome of the EORTC 20981 Phase III randomized intergroup study. Journal of Clinical Oncology. 2010;**28**(17):2853-8.

31. National Institute for Health and Clinical Excellence. Rituximab for the treatment of relapsed or refractory stage III or IV follicular non-Hodgkin's lymphoma (review of technology appraisal guidance 37):TA137. London: NICE; 2008 [accessed 2010 25th August]; Available from: http://guidance.nice.org.uk/TA137/Guidance/pdf/English.
32. Roche Ltd. Rituximab - Clinical study report (M018264)-No.10347952010
33. Davis TA, Grillo-Lopez AJ, White CA, McLaughlin P, Czuczman MS, Link BK, et al. Rituximab anti-CD20 monoclonal antibody therapy in non-Hodgkin's lymphoma: safety and efficacy of re-treatment. Journal of Clinical Oncology. 2000;**18**(17):3135-43.
34. Igarashi T, Ohtsu T, Fujii H, Sasaki Y, Morishima Y, Ogura M, et al. Re-treatment of relapsed indolent B-cell lymphoma with rituximab. Int J Hematol. 2001;**73**(2):213-21.
35. Johnston A, Bouafia-Sauvy F, Broussais-Guillaumot F, Michallet AS, Traulle C, Salles G, et al. Retreatment with rituximab in 178 patients with relapsed and refractory B-cell lymphomas: a single institution case control study. Leuk Lymphoma. 2010;**51**(3):399-405.
36. Coiffier B, Bouaffia F, Thieblemont C, Hequet O, P A, Dumontet C, et al. Rituximab re-treatment in B-cell lymphomapatients: efficacy and toxicity in 59 patients treated in one center. Blood. 2002;**100**:Abstract 1390.
37. Cheung MC, Haynes AE, R M, A S, Imrie KR. Rituximab in lymphoma: a systematic review and consensus practice guideline from Cancer Care Ontario. Cancer Treat Rev. 2007;**33**(2):161-76.
38. Electronic Medicines Compendium. Browse medicines. 2010 [accessed 2010 September 2010]; Available from: http://www.medicines.org.uk/EMC/medicine/14048/SPC/Zevalin+1.6+mg+ml%2c+Kit+for+radiopharmaceutical+preparation+for+infusion/.
39. IMS Health. Hospital Pharmacy Audit.2010
40. Moher D, Chalmers T, Dook D, Eastwood S, Olkin I, Drummond R, et al. Improving the quality of reports of meta-analysis of randomized controlled trials: the Quorom statement. Lancet. 1999;**354**:1896-900.
41. Cheson BD, Horning SJ, Coiffier B, Shipp MA, Fisher RI, Connors JM, et al. Report of an international workshop to standardize response criteria for non-Hodgkin's lymphomas. NCI Sponsored International Working Group. Journal of Clinical Oncology. 1999 Apr;**17**(4):1244.
42. European Organisation for Research and Treatment of Cancer. EORTC QLQ-C30. 2010 [accessed 2010 September 1st]; Available from: http://groups.eortc.be/qol/questionnaires_qlqc30.htm.
43. Functional Assessment of Chronic Illness Therapy. FACIT. 2010 [accessed 2010 1st September]; Available from: http://www.facit.org/qview/qlist.aspx.
44. Centre for Reviews and Dissemination. Systematic Reviews: CRD's guidance for undertaking reviews in healthcare2008: Available from: http://www.york.ac.uk/inst/crd/pdf/Systematic_Reviews.pdf.
45. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER). Guidance for industry, clinical trial endpoints for the approval of cancer drugs and biologics. 2005.
46. Floriana I, Rotmensz N, Albertazzi E, Torri V, De Rosa M, Tomino C, et al. Approaches to interim analysis of cancer randomised clinical trials with time to event points: a survey from the Italian National Monitoring Centre for Clinical Trials. Trials. 2008;**9**(46).
47. Proschan MA, Lan KK, Wittes JT. Statisical Monitoring of Clinical Trials: a Unified Approach. New York: Springer; 2006. p. 99-111.

48. Pettengell R, Donatti C, Hoskin P, Poynton C, Kettle PG, Hancock B, et al. The impact of follicular lymphoma on health-related quality of life. Ann Oncol. 2008;**19**:570-76.
49. Wild D. Utility values in follicular lymphoma - Oxford Outcomes. Unpublished report by Oxford Outcomes prepared for Roche UK: Oxford Outcomes2005
50. Wild D, Walker M, Pettengell R, Lewis G. Utility elicitation in patients with follicular lymphoma Value in Health. 2006;**9**(Abstract PCN62).
51. Department of Health. NHS Reference Costs 2008-20092010: Available from: http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_111591 (Accessed: 25/08/2010).
52. British National Formulary. BNF 59. 2010; Available from: http://www.bnf.org/bnf/ (Accessed: 06/07/2010).
53. Curtis L. Unit Costs of Health and Social Care 2009. Personal Social Services Research Unit, University of Kent; 2009; Available from: http://www.pssru.ac.uk/uc/uc.htm.
54. National Institute for Health and Clinical Excellence. Improving supportive and palliative care for adults with cancer. 2004 [accessed 2010 October]; Available from: http://www.nice.org.uk/nicemedia/live/10893/28816/28816.pdf.
55. Johnson PW, Rohatiner AZ, Whelan JS, Price CG, Love S, Lim J, et al. Patterns of survival in patients with recurrent follicular lymphoma: a 20-year study from a single center. Journal of Clinical Oncology. 1995;**13**(1):140-7.
56. Rohatiner AZS, Lister TA. The clinical course of follicular lymphoma. Bailliere's Best Practice in Clinical Haematology. 2005;**18**(1):1-10.
57. Bassler D, Briel M, Montori VM, Lane M, Glasziou P, Zhou Q, et al. Stopping randomized trials early for benefit and estimation of treatment effects: systematic review and meta-regression analysis. JAMA. 2010 Mar 24;**303**(12):1180-7.
58. Trotta F, Apolone G, Garattini S, Tafuri G. Stopping a trial early in oncology: for patients or for industry? Ann Oncol. 2008;**19**:1347-53.
59. Pocock SJ. When to stop a clinical trial. BMJ. 1992;**305**:235-40.
60. Ioannidis JPA, Karassa FB. The need to consider the wider agenda in systematic reviews and meta-analyses: breadth, timing, and depth of the evidence. BMJ. 2010;**341**.

8 APPENDICES

Appendix 1 Quality assessment of included trial

Table 29 Quality assessment of PRIMA trial

Study question	How is the question addressed in the study?	Grade (yes/no/not clear/N/A)	ERG comment
Was randomisation carried out appropriately?	Centralised, stratified block randomisation procedure.	Yes	This was appropriate
Was the concealment of treatment allocation adequate?	Randomisation of eligible patients was performed centrally by fax from the GELA randomization center (GELARC) at Hôpital Saint Louis–Centre Hayem. The random allocation sequence was generated by an SAS programmer according to the specifications of a biostatistician. The SAS database that was imported in the GELARC randomisation tool was not readable. Thus, neither the physicians nor the randomisation assistants had access to the random allocation sequence, which was kept by the biostatistics department.	Yes	Agree that this was an adequate system
Were the groups similar at the outset of the study in terms of prognostic factors, for example, severity of disease?	The treatment groups were well balanced with respect to follicular lymphoma international prognostic index (FLIPI) scores (see section 5.3.4, Table 14)	Yes	The groups appeared to be similar at induction and maintenance phases
Were the care providers, participants and outcome assessors blind to treatment allocation? If any of these people were not blinded, what might be the likely impact on the risk of bias (for each outcome)?	PRIMA was an open-label study, therefore, it is likely that the aforementioned parties were aware of treatment allocation. However, the assessment of follicular lymphoma post-treatment is very objective and it is therefore unlikely that this will have biased results. In addition, an IRC comprising three hemato-oncologists and seven radiologists (including two adjudicators) assessed all patients randomized in the maintenance/observation phase in a blinded manner for response and progression based on computed tomography (CT) scans and reports of pertinent clinical findings (including physical examination and laboratory results) according to the IRC Charter.	No	Open-label RCTs are common in this disease area. The risk of bias was minimised by the use of the IRC. The risk of bias in terms of the patient group is likely to be minimal given the objective nature of response assessment. There is a possibility of bias QoL responses but this appeared not to be the case in the PRIMA trial.
Were there any unexpected imbalances in drop-outs between groups? If so, were they explained or adjusted for?	263 patients (26%) discontinued during the maintenance/observation phase (Table 10). More patients in the observation arm than in the RTX arm withdrew from the study (162 patients vs 101 patients; 32% vs 20%). See section 5.3.1.2.18 for details.	No	All withdrawals were clearly explained.
Is there any evidence to suggest that the authors measured more outcomes than they reported?	All pre-defined primary and secondary outcomes have been reported.	No	All outcomes specified in the protocol appear to be reported.
Did the analysis include an intention-to-treat analysis? If so, was this appropriate and were appropriate methods used to account for missing data?	As detailed in section 5.3.1.2.16. Efficacy and economic analyses are subsequently presented for the intention-to-treat population. This was an appropriate approach in order to preserve the randomisation scheme and avoid selection bias. A sensitivity analysis of investigator-assessed PFS was performed to account for missing data.	(i) Yes (ii) Yes	ITT included using appropriate methods.
Centre for Reviews and Dissemination (2008) Systematic reviews. CRD's guidance for undertaking reviews in health care. York: Centre for Reviews and Dissemination[44]			

Appendix 2

Definitions used in key trial

Table 30 Definitions used in PRIMA trial

Terms used	Definition
Complete response (CR)	•disappearance of all detectable clinical and radiographic evidence of disease as well as disease-related symptoms if present before therapy, and normalization of NHL-related biochemical abnormalities (eg, LDH) definitely assignable to NHL • all lymph nodes and nodal masses must have regressed to normal size (≤ 1.5 cm for nodes which were ≥ 1.5 cm before therapy). Nodes that were 1.1–1.5 cm prior to therapy must have decreased to < 1 cm in their greatest transverse diameter after therapy, or by more than 75% in the SPD • spleen and other organs (e.g. liver, kidneys) considered enlarged before therapy must have decreased in size and not be palpable on physical examination; • bone marrow, if positive at baseline, must be negative.
Complete response unconfirmed (CRu)	The first and third criteria above were met, but with one or more of the following features: • a residual lymph node mass greater than 1.5 cm in its greatest transverse diameter that regressed by more than 75% in the SPD. Individual nodes that were previously confluent must have regressed by more than 75% in their SPD compared with the size of the original mass • indeterminate bone marrow
Event-free survival (EFS)	Measured from the date of randomisation to the maintenance/observation phase to the date of first documented progression, relapse, initiation of a new anti-lymphoma treatment (CTX, radiotherapy, radioimmunotherapy, immunotherapy) or death from any cause. Patients who did not experience an event at the time of the analysis were censored at the date of the last clinical examination or imaging performed by the investigator.
Overall response rate (ORR)	Assessed at the end of the two-year maintenance/observation phase by the investigator. Assessment of response was based on the 1999 International Workshop criteria for evaluation of response in non-Hodgkin's lymphomas[41]
Overall survival (OS)	Measured from the date of randomisation to the maintenance/observation phase to the date of death, regardless of cause. Patients who had not died at the time of the analysis were censored at the date the patients were last known to be alive.
Partial response (PR)	≥ 50% decrease in SPD of the six largest dominant nodes or nodal masses. These nodes or masses should be selected as follows: (a) they should be clearly measurable in at least two perpendicular dimensions, (b) they should be from as disparate regions of the body as possible, and (c) they should include mediastinal and retroperitoneal areas of disease whenever these sites are involved • no increase in the size of the other nodes, liver, or spleen • splenic and hepatic nodules must regress by at least 50% in the SPD • with the exception of splenic and hepatic nodules, involvement of other organs is considered assessable and not measurable disease • no new sites of disease
Progression-free survival (PFS)	Measured from the day of randomisation to the maintenance/observation phase to the date of first documented disease progression, relapse, or death from any cause. Assessment of progression and relapse was based on the criteria for evaluation of response in NHL[41]
Progressive disease	• ≥ 50% increase from nadir in the SPD of any previously identified abnormal node • appearance of any new lesion during or at the end of therapy

Terms used	Definition
Quality of life	Measurements were made using the FACT-G and QLQ-C30 questionnaires collected over time, and were compared for patients receiving RTX maintenance and for those in the observation arm
Stable disease (SD)	Less than a PR but not progressive disease
Time to next anti-lymphoma treatment	Measured from the date of randomisation to the maintenance/observation phase to the date of first documented administration of any new anti-lymphoma treatment (CTX, radiotherapy, radioimmunotherapy, immunotherapy). Patients who did not experience an event at the time of the analysis were censored at the date of the last clinical examination or imaging performed by the investigator.
Time to next CTX treatment	Measured from the date of randomisation to the maintenance/observation phase to the date of first documented administration of new CTX or new cytotoxic agent. For any given patient, the TTNCT may be the same as the TTNLT. Patients who did not experience an event at the time of the analysis were censored at the date of the last clinical examination or imaging performed by the investigator.
Transformation rate at first progression	Restricted to patients with a biopsy at first progression. This parameter was defined by the appearance of diffuse areas of large lymphoma cells within a tumor site. For this purpose, a biopsy or a cytological examination was obtained at progression, where possible, and made available for central pathological review

SPD= sum of the products of the greatest diameters

Appendix 3

Summary of secondary analyses

Table 31 Summary of secondary analyses

Analysis	Page in MS	Manufacturer's summary as presented in MS
Investigator-Assessed PFS in the Per Protocol Population (PPP)	104	The analysis of PFS based on the PPP (410 patients in the observation arm, and 434 patients in the RTX arm) was consistent with the analysis based on the MITT population. In the PPP, 152 patients on observation and 80 patients on RTX maintenance treatment (37.1% vs 18.4%) experienced a PFS event. The PP analysis confirmed the results from the primary ITT analysis that RTX maintenance treatment significantly reduced the risk of experiencing a PFS event. In the PP analysis, the risk of experiencing a PFS event was reduced by 55% compared with observation (stratified HR 0.45, 95% CI [0.34;0.59], p < 0.0001, stratified log-rank test). The estimated 25th percentile times to progression were 482 days (15.8 months) for patients on observation and 1096 days (36.0 months) for patients on RTX maintenance in the per protocol population. Results of the unstratified sensitivity analysis of PFS in the per protocol population were similar to results of the stratified analysis. The corresponding Kaplan–Meier curve and summary of event-free rates over time are available on request.
IRC-assessed PFS	108	Of 1018 patients in the MITT population, [redacted] patients were assessed for disease progression by the IRC. An IRC review was missing for [redacted] patients ([redacted] patients in the observation arm, and [redacted] patients in the RTX arm). The reasons why patients were missing from the IRC analysis included: additional informed consent for CT scan collection not obtained ([redacted] patients: [redacted] patients in the observation arm and [redacted] patients in the RTX arm); no radiological evidence of disease at baseline and subsequent scans not evaluable ([redacted] patients: [redacted] observation and [redacted] RTX); missing baseline scans ([redacted] patients: [redacted] observation and [redacted] RTX); baseline scan not readable ([redacted] patients: [redacted] observation and [redacted] RTX); and no post-baseline scan available ([redacted] patient in the RTX arm). The baseline characteristics for those patients with no IRC assessment were similar between the two arms and similar to those of the MITT population. For the primary analysis of IRC-assessed PFS, patients without an independent assessment were censored at day 1. At the time of the analysis, [redacted]/1018 patients ([redacted]%) had experienced disease progression according to the IRC's assessments or death). More patients in the observation arm experienced a PFS event than in the RTX arm ([redacted] vs [redacted] patients, [redacted] vs [redacted] Disease progression/relapse was recorded for 154 patients in the observation arm and [redacted] patients in the RTX arm ([redacted] There were [redacted] deaths in the observation arm and [redacted] deaths in the RTX arm before IRC-assessed progression/relapse. The risk of experiencing IRC-confirmed disease progression/relapse or death was reduced by 46% for patients receiving RTX maintenance compared to those on observation (stratified HR 0.54, 95% CI [0.42;0.70]). Maintenance therapy with RTX significantly prolonged IRC-assessed PFS as compared to observation (p < 0.0001, stratified log-rank test)_The Kaplan–Meier estimated median PFS was [redacted] days ([redacted] months) for the observation arm and [redacted] days ([redacted] months) for the RTX arm. It is worth mentioning that the median is not a good measure for the treatment benefit if it is only reached at the tail end of the Kaplan–Meier curve, as is the case for the IRC-assessed PFS presented here. The median is then always highly dependent on the very few patients still at risk at the time the median is reached (in this case, less than [redacted] patients were at risk in the RTX arm). The more robust measure is the hazard ratio as it takes into consideration the entire observation period when estimating the treatment difference. This also explains why there was a comparable treatment benefit in terms of the hazard ratio between the investigators' and IRC's assessments of PFS, although in the investigator-based PFS analysis

		[redacted]On the basis of Kaplan–Meier estimates, [redacted] of patients in the RTX arm and [redacted] of patients in the observation arm were alive and progression-free at one year. At two years, [redacted] (95% CI [redacted]) of patients in the RTX arm and [redacted]% (95% CI [redacted]) of patients in the observation arm were progression-free. Results of the analysis of IRC-assessed PFS without stratification by induction treatment or response to induction treatment were similar to those of the stratified analysis (non-stratified HR [redacted], 95% CI [redacted], [redacted], non-stratified log-rank test) (
IRC-assessed PFS in the PPP	112	IRC-assessed PFS was also analyzed in the per protocol population ([redacted] patients in the observation arm, and [redacted] patients in the RTX arm), and the results were consistent with the MITT analysis of IRC-assessed PFS. In the PPP, [redacted] patients in the observation arm and [redacted] patients in the RTX maintenance arm ([redacted]% vs [redacted]%) experienced a PFS event. The risk of experiencing a PFS event was reduced by [redacted]% for patients receiving RTX maintenance treatment compared with those on observation (stratified HR [redacted], 95% CI [redacted]). The per protocol analysis confirmed the results from the MITT IRC-assessed PFS analysis that RTX maintenance therapy significantly prolonged PFS as compared to observation ([redacted], stratified log-rank test). The Kaplan–Meier estimated median PFS was [redacted] days ([redacted] months) in the observation arm and [redacted] days ([redacted] months) in the RTX arm. Note again that only very few patients were at risk at the time the median was reached in the RTX arm (less than [redacted] patients in the RTX arm were at risk) and therefore the median in this case is not a robust measure of the observed treatment difference. Results of a per protocol sensitivity analysis without stratification were similar to the results of the stratified analysis.
IRC-assessed PFS subgroups analysis	113	As for investigator-assessed PFS, the potential impact of baseline demographics and prognostic factors on the treatment effect as assessed by the IRC was analyzed for the following subgroups: age (≥ 60 years, < 60 years), gender (male, female), FLIPI score (≤ 1, 2, ≥ 3), induction treatment (R-CHOP, R-CVP, R-FCM), and response to induction treatment (CR/CRu, PR). Hazard ratios for PFS with 95% confidence intervals (observation vs RTX) for the prespecified patient subgroups are shown on the Forrest plot The risk of disease progression or death was reduced in the RTX arm compared to the observation arm in all of the subgroups tested [redacted] Overall, the results of the IRC-assessed PFS subgroup analyses are consistent with the primary analysis of IRC-assessed PFS in the MITT population, and with the investigator-assessed PFS analyses (Section 5.5.2.1.1). A multiple Cox regression was performed with the covariates age (≥ 60 years,< 60 years), gender (male, female), FLIPI score (≤ 1, 2, ≥ 3), induction treatment (R-CHOP, R-CVP, R-FCM), and response to induction treatment (CR/CRu, PR). The hazard ratio for RTX maintenance versus observation adjusted for all the covariates and its 95% confidence interval (multivariate adjusted HR [redacted] 95% CI [redacted] was similar to the result obtained from the unadjusted analysis (unadjusted HR 0.54, 95% CI [0.42;0.70]). An additional Cox analysis applying a stepwise backward selection procedure was performed. The hazard ratio adjusted for the covariates and its 95% confidence interval (stepwise backward selection model, multivariate adjusted HR [redacted] 95% CI [redacted] was similar to the result obtained from the unadjusted analysis. The results of univariate Cox regression analyses confirmed that the treatment difference adjusted for each covariate favored maintenance therapy with RTX ([redacted]) and was consistent with the unadjusted Cox regression (unadjusted HR 0.54, 95% CI [0.42;0.70]).
Comparison of investigator and IRC assessments	115	Data for 887 patients from the MITT population were reviewed for disease response and progression by the IRC as well as by the investigators: 447 patients in the observation arm and 440 patients in the RTX arm (87% vs 87%). An independent assessment was missing for 131 patients (13% overall; see Section 5.5.2.1.2). In the dataset of 887 patients for whom

		both reviews were available, there were [redacted] PFS events according to the investigators' assessments and [redacted] PFS events according to the IRC's assessments. Concordance between the investigators' and IRC's assessments of disease progression was analyzed in terms of whether or not disease progression occurred and the time at which disease progression occurred. The concordance/discordance between the investigators' and IRC's assessments of disease progression is summarized in Table 27. Of the [redacted] patients in the observation arm with both reviews, investigator and IRC agreed that an event (progression or death) occurred in [redacted] patients and did not occur in [redacted] patients, resulting in a concordance rate of [redacted] for the observation arm. Similar results were observed in the RTX arm: the investigators and the IRC agreed that an event (progression or death) occurred in [redacted] patients and did not occur in [redacted] patients, resulting in a concordance rate of [redacted]% for the RTX maintenance arm. Concordance/discordance for the investigators' and IRC's assessments was also analyzed in terms of the timing of disease progression. This analysis was based on patients who were considered to have progressed according to the investigator, the IRC, or both). Concordant timing was defined as disease progression within 30 days according to the IRC and the investigator. Among the [redacted] patients with disease progression in the observation arm according to both the IRC and investigator, [redacted] events ([redacted]%) had concordant timing, [redacted] events ([redacted]%) occurred earlier according to the IRC, and [redacted] events ([redacted]%) occurred later according to the IRC ([redacted] Among the [redacted] patients with disease progression in the RTX arm according to both the IRC and investigator, [redacted] events ([redacted]%) had concordant timing, [redacted] events ([redacted]%) occurred earlier according to the IRC, and [redacted] events ([redacted]%) occurred later according to the IRC ([redacted]). For both arms, discordance in the timing of disease progression was mostly due to progression events that occurred earlier based on IRC assessments compared with the investigators' assessments ([redacted] in the observation arm and [redacted] in the RTX arm; [redacted]% vs [redacted]%).
Additional robustness analyses of PFS	119	The robustness of the PFS result was assessed by performing sensitivity analyses. These analyses confirmed the significant results of the primary analyses of PFS based on the investigators' assessments and the IRC's assessments and indicated that maintenance therapy with RTX significantly prolonged PFS and reduced the risk of a PFS event.

Appendix 4

Summary of safety outcomes in the PRIMA[32] trial

Table 32 Summary of safety outcomes

Analysis	Page in MS	Manufacturer's summary as presented in the MS
Toxicities	162	On the basis of the checklist of prespecified toxicities in the CRF, 97% of patients in the RTX arm and 90% of patients in the observation arm experienced at least one toxicity during the maintenance/observation phase (Table 56). Toxicities that were more prevalent in the RTX arm than the observation arm included constitutional symptoms, decreases in leukocytes, neutrophils, or hemoglobin, infection with normal neutrophils, gastrointestinal disorders, increases in transaminases (AST/ALT), and pulmonary disorders. The vast majority (> 90%) of these toxicities were mild/moderate in severity (ie, Grade 1/2 events). Note that an additional 5% of patients in the observation arm and 3% of patients in the RTX arm had toxicities recorded under the 'other' category, which is not present.
AEs	164	The proportion of patients who experienced at least one AE (including Grade 3–5 toxicities, Grade 2–5 infections, and SAEs) during the maintenance/observation phase was higher in the RTX arm than in the observation arm (52% vs 35%). This difference was mainly due to infections and infestations (37% of patients in the RTX arm vs 22% of patients in the observation arm). A total of 728 AEs were reported. The most common categories of AEs were infections and infestations (mainly bronchitis), neoplasms (mainly basal cell carcinoma), and blood and lymphatic system disorders (mainly neutropenia). The incidence of other categories of AEs was low (< 4%) and similar in the two study arms More AEs were reported in the RTX arm compared with the observation arm (459 vs 269 events). AEs that occurred with a higher incidence (≥ 2% difference) in the RTX arm compared with the observation arm were bronchitis (9% vs 5%), upper respiratory tract infection (5% vs 2%), sinusitis (4% vs 2%), and neutropenia (4% vs <1%). The incidence of all other AEs was low and comparable between the two study arms. The majority of AEs were Grade 2 in severity (165/269 AEs [61%] in the observation arm; 291/459 AEs [63%] in the RTX arm), and the majority (85%) of those events were Grade 2 infections (144 events in the observation arm, and 248 events in the RTX arm). A total of 195 patients recorded Grade 3 or 4 AEs (see Section 5.9.2.4.3). There were five Grade 5 (fatal) AEs.
AEs over time	165	The proportion of patients reporting AEs was slightly higher in the RTX arm than in the observation arm at almost all visits, but there were no apparent trends for increasing incidence with time in either arm with subsequent cycles of treatment/observation (Table 59).
AEs by cumulative TX dose	166	The number of Grade 2–5 infections appears to increase with cumulative RTX dose. However, comparison with the number (percentage) of infections by treatment cycle shows that the number of infections recorded at each cycle did not increase over time (ie, the number of infections occurring in cycles 10, 11, or 12 was no greater than the number occurring in early cycles [eg, cycles 1, 2, or 3]). Therefore, the higher incidence of infections seen with increasing cumulative RTX dose simply reflects the link between cumulative dose and longer overall observation time rather than the cumulative RTX dose per se.
AEs:treatment-related	168	AEs which were reported by the investigator as having a remote, possible, or probable relationship to trial treatment were reported for 9% of patients in the observation arm and 29% of patients in the RTX arm. Overall, 60/269 events (22%) in the observation arm and 229/459 events (50%) in the RTX arm were considered to be related to study treatment. AEs which were most commonly considered to be related to trial treatment included infections and infestations, and blood and lymphatic disorders (Table 61). The occurrence of 'treatment-related' AEs in the observation arm despite no treatment (or lymphoma) being administered during the maintenance/observation phase probably reflects toxicities associated with the induction therapy being carried over into the maintenance/observation phase (a similar effect was probably also present in the RTX arm).

Analysis	Page in MS	Manufacturer's summary as presented in the MS
AEs: age-related		AEs occurring in patients aged under 65 years old, from 65 to 74 years old inclusive, and 75 years old and over, were analyzed for US regulatory purposes (Table 62). Although the overall incidence of AEs appeared to increase with age in the observation arm, this was not apparent in the RTX arm. The overall incidence of infections and infestations, blood and lymphatic system disorders, and neutropenia AEs also showed no clear increase with age in either arm, bearing in mind the low numbers of patients in the ≥ 75 years age group. Infections were the most commonly occurring AEs in all three age categories. For patients aged under 65 years, the incidence of AEs was higher in the RTX arm compared with the observation arm (54% vs 32%). This difference was mainly due to a higher incidence of infections (mainly bronchitis, upper respiratory tract infection, and sinusitis) and blood and lymphatic system disorders (mainly neutropenia) in the RTX arm. For patients aged 65–74 years, the overall incidence of AEs as well as the incidence of infection AEs was balanced between the two study arms (34% in the RTX arm vs 29% in the observation arm). In the ≥ 75 year age group, the incidence of AEs was similar between the two study arms but the incidence of infection AEs was higher in the RTX arm than in the observation arm (35% vs 17%). Comparing incidences of Grade 3, 4, and 5 AEs, and infection and infestation AEs in particular, across the different age groups, the overall incidence again appeared to increase with age in the observation arm but only slightly in the RTX arm (Table 63). There was no apparent difference in incidence of AEs in the different age groups. In particular, patients aged > 60 years old did not appear to experience more infection or hematological AEs than younger patients.
AEs: Grade 3 or 4	174	More patients in the RTX arm than in the observation arm experienced at least one Grade 3 or 4 adverse event (24% vs 16%). This difference was mainly due to a higher incidence of Grade 3 or 4 AEs of blood and lymphatic system disorders (mainly neutropenia) and infections in the RTX arm than in the observation arm. Note that Grade 5 AEs (those with a fatal outcome) were analyzed separately (see Section 5.9.2.5). Grade 3/4 AEs occurring with an incidence of 1% or higher in either arm are summarized in Table 64. Grade 3/4 AEs that occurred with a higher incidence (≥ 2%) in the RTX arm compared with the observation arm was neutropenia (4% vs < 1%). Grade 3/4 leukopenia was also more common in the RTX arm than in the observation arm (2% vs <1%). The incidence of all other Grade 3/4 AEs was low (<1%) and comparable between the two study arms.
Deaths	176	At the time of clinical cut-off (January 14, 2009), a total of 31 patients had died during the active maintenance/observation phase or during follow-up (Note that the number of deaths based on the MITT population was 34 the difference being accounted for by the deaths of three patients (patients 10140/1004, 20334/1001, and 60143/1002) randomized to the RTX arm but withdrawn from the study prior to receiving treatment.) The number of deaths was higher in the observation arm than in the rituximab arm (18 patients vs 13 patients). The most common cause of death was disease progression (lymphoma), which accounted for 12 deaths in the observation arm and 10 deaths in the RTX arm. The incidence of non-lymphoma deaths was higher in the observation arm than in the RTX arm (six patients vs three patients). Five of the nine non-lymphoma deaths were considered to be outcome of AEs. Two fatal AEs in the observation arm were a result of neoplasms: leukemia considered as possibly related to trial treatment (patient 10222/1008) and metastatic neoplasm considered to be treatment-unrelated (patient 21011/1013). The three recorded fatal AEs in the RTX arm resulted from a treatment-unrelated disorder (unknown/unevaluable event; patient 20439/1003), hepatitis B considered to be probably treatment-related (patient 20111/1016), and pulmonary haemorrhage considered to be treatment-unrelated (patient 20731/1008). The remaining four deaths (not due to lymphoma or to AEs) were all in the observation arm and were due to acute myeloid leukemia (patient 71501/1013), coronary artery disease (patient 10307/1013), myelodysplastic syndrome (patient 40346/1009), and sepsis (patient 60113/1004).

Analysis	Page in MS	Manufacturer's summary as presented in the MS
SAEs	178	A total of 193 SAEs were reported for 158 patients (63 patients [12%] in the observation arm, and 95 patients [19%] in the RTX arm) during the maintenance/observation phase Note that all SAEs occurred with a incidence of less than 1% in both arms. The most common class of SAEs overall was neoplasms (39 events overall affecting 37 patients), including basal cell carcinoma (two patients in the observation arm vs four patients in the RTX arm), colon cancer (three patients in the RTX arm) and breast cancer (two patients in the RTX arm). The most common class of SAE in the RTX arm was infections and infestations (25 patients [5%] vs six patients [1%] in the observation arm). In the RTX arm, three patients had SAEs of pneumonia, two patients had diverticulitis, and two patients had hepatitis B (see Section 5.9.2.9.2). In the observation arm, three patients had SAEs of urinary tract infections. Other serious infections were reported by only one patient in each case. Serious cardiac disorders were reported for two patients in the observation arm compared with 11 patients in the RTX arm (see Section 5.9.2.9.4).
AEs leading to treatment discontinuation	179	A total of 27 patients discontinued maintenance treatment/observation as a result of adverse events (eight patients [2%] in the observation arm, and 19 patients [4%] in the RTX arm). The most common AEs that led to treatment discontinuation were neoplasms, which accounted for the withdrawal of six patients in the observation arm and five patients in the RTX arm. Four patients in the RTX arm were withdrawn as a result of infections: hepatitis B (two patients), endocarditis, and mycobacterial infection. One case of hepatitis B was considered to be unrelated to trial treatment, and the other three infections were considered as being probably treatment-related. Five patients discontinued treatment after becoming pregnant (Section 5.9.2.10).
AEs leading to treatment discontinuation	180	A total of 30 patients had their dosing of RTX interrupted or modified as a result of an adverse event. The most common reasons for interrupting the dose schedule or for modifying the RTX dose were infections and infestations (12 patients) including three bronchitis events and two upper respiratory tract infections, and blood and lymphatic disorders (nine patients) including seven neutropenia events and five leukopenia events.
AEs of special interest		
Infusion-related reactions	181	Adverse events occurring within one day of a RTX cycle or an observation visit were analyzed to determine the pattern of potential infusion-related reactions. More AEs were reported in the RTX arm than in the observation arm within one day after a treatment cycle/observation visit (74 events in 61 patients [12%] vs 61 events in 46 patients [9%]). The majority of these AEs were infections (mainly upper respiratory tract infection and bronchitis). Typical RTX infusion-related AEs, such as chills, pyrexia, nausea, and vomiting, were not reported in the RTX arm (only one administration site disorder [mucosal inflammation] was reported), indicating that if they had occurred then they were less than Grade 3 in severity. The view that infusion-related reactions (if they occurred) were mainly Grade 1 or 2 in severity during the maintenance/observation phase is supported by the finding that the checklist toxicity 'constitutional symptoms' was reported in 155 patients in the observation arm (31%) and 203 patients (41%) in the RTX arm throughout the maintenance/observation phase). Most of these were Grade 1 or 2 in severity. A low incidence of severe infusion-related reactions was expected given that patients were previously exposed to RTX during the induction phase and were routinely premedicated with an analgesic/antipyretic and an antihistamine before each infusion of RTX, Tumor lysis syndrome was also not expected to occur during the maintenance/observation phase since this complication is generally associated with initial treatment of patients with bulky disease. No cases were reported during the maintenance/observation phase (although three cases were reported during the induction phase).

Analysis	Page in MS	Manufacturer's summary as presented in the MS
Infections and infestations	183	Infections were not collected as a single category on the checklist toxicity CRF but according to neutrophil count More patients in the RTX arm than in the observation arm (192 patients vs 127 patients, 38% vs 25%) had an infection with normal neutrophil count. In addition, 11 patients (2%) in the RTX arm had an infection with Grade 3/4 neutropenia (nine patients (2%) with Grade 2 infections, and two patients (<1%) with Grade 3 infections) compared with two patients (< 1%) in the observation arm (both Grade 2). One additional patient in each arm had febrile neutropenia (Grade 3). Infections (Grade ≥ 2) were the most common class of AEs recorded, and the incidence was higher in the RTX arm than in the observation arm (184 patients vs 114 patients, 37% vs 22%). However, most infections were Grade 2 in severity—the incidence of Grade 3 or 4 infections was only 4% in the RTX arm and 1% in the observation arm. Similarly, infection SAEs occurred in 25 patients (5%) in the RTX arm and in six patients (1%) in the observation arm. The proportion of patients with an infection AE that was considered to be treatment-related was 6% in the observation arm (30 patients) and 21% in the RTX arm (107 patients). Four patients in the RTX arm discontinued treatment and a further 12 patients had their RTX dosing modified or interrupted as a result of infection AEs. One patient in the RTX arm died of hepatitis B infection (patient 20111/1016—this event was coded under hepatobiliary AEs rather than infections and infestations), and one patient in the observation arm died of sepsis (patient 60113/1004). A manual search of AE preferred terms to see if any other infections were included in other categories only revealed six cases of conjunctivitis (two cases in the observation arm, and four cases in the RTX arm)—all cases were Grade 2 in severity and resolved without sequelae. Most infections had no causal organism documented. Of those infection AEs with an indentified organism, the most common pathogens were viral (22 patients [4%] in the observation arm and 28 patients [6%] in the RTX arm), bacterial (11 patients [2%] vs 23 patients [5%]and fungal (three patients [<1%] vs nine patients [2%]).
Hepatitis B	184	Three patients had hepatitis B reported during the maintenance/observation phase of the study. Two of these patients had reactivation of hepatitis B infection, and one patient's past hepatitis B status was unknown. along with patients with AEs of hepatitis B reported during the induction phase. Four additional patients had AEs reported that could potentially have been related to hepatitis B. Two of these cases were clearly due to other causes: patient 10139/1005 (cytolytic hepatitis reported during the induction phase due to infection of a biliary stent) and patient 71101/1059 (ascites reported while the patient was on observation—the patient had a known history of hepatic cirrhosis). The third patient (patient 73001/1126) developed fulminant hepatitis after seven cycles of R-CVP. The cause was uncertain, but hepatitis B serology was negative (as was other viral serology, including hepatitis A and hepatitis C). This event occurred after removal of an intravenous port for suspected (bacterial) infection, and the hepatitis resolved without sequelae. The fourth case (patient 10109/1006: Grade 4 cytolytic hepatitis reported on day 787) was not reported as an SAE and little information is available. However, the cytolytic hepatitis was reported as an unrelated adverse event 11 days after the diagnosis of progressive disease. Overall, therefore, the incidence of hepatitis B was less than 1% (six patients) in the study, but importantly three of the six patients had a fatal outcome. Interestingly, at least three patients with a known history of hepatitis B infection entered the study and did not develop hepatitis B reactivation during treatment. Patient 20731/1025 received prophylactic lamivudine and completed R-CHOP induction treatment followed by observation without reactivation; patient 41030/31031 received prophylactic lamivudine and completed R-CHOP induction treatment and RTX maintenance without reactivation; and patient 10236/1002 received no prohylaxis and completed R-CHOP induction treatment without reactivation (the patient's lymphoma progressed during the observation phase).

Analysis	Page in MS	Manufacturer's summary as presented in the MS
Progressive multifocal leukoencephalopathy	187	Two cases of progressive multifocal leukoencephalopathy (PML) were reported in the course of this study and are described here for completion, although both cases were reported after the second clinical cut-off date for the updated analysis (June 30, 2009). One case was reported as an SAE after the second clinical cut-off date (patient 10109/1015). This patient was in the RTX maintenance arm of the study and developed PML in the context of disease progression after extensive subsequent therapy, including R-ICE (RTX, ifosfamide, cytarabine, and etoposide), high-dose chemotherapy with stem cell rescue, and investigational therapy targeting CD19 (B-cells). The patient developed neurological symptoms after the investigational therapy and subsequently died of progressive disease and PML. With the second case (patient 73001/1117), PML was not reported as an AE within the PRIMA trial but was described as the cause of death (in August 2009). This patient, who was in the observation arm of the study, developed progressive disease while on observation and received treatment with RTX and Apomab (an investigational antibody directed against human death receptor 5 [DR5; TRAIL-R2; TNFRSF10B]). Eleven months later, the patient died due to PML. The investigator considered the patient's death to be due to the toxicity of the subsequent therapy.
Blood and lymphatic system disorders	187	As expected, blood and lymphatic system AEs were reported for more patients in the RTX arm compared with the observation arm (34 events in 26 patients [5%] vs nine events in seven patients [1%]). The majority of these events were neutropenia (19 patients in the RTX arm and five patients in the observation arm). Grade 3 or 4 neutropenia were recorded for 18 patients in the RTX arm compared with five patients in the observation arm. Two cases of febrile neutropenia (one patient in each study arm) and two cases of neutropenia (one patient in each study arm) were also reported as SAEs. The two patients who developed SAEs of neutropenia/febrile neutropenia in the observation arm are of note because the neutropenia appeared to occur after some delay. Patient 10105/1037 developed Grade 3 febrile neutropenia 123 days after the last dose of RTX, and patient 20334/1008 developed Grade 4 neutropenia 148 days after the last dose of RTX. Seven patients in the RTX arm had their RTX dosing modified or interrupted as a result of neutropenia. Three of these seven patients received granulocyte-colony stimulating factor as a treatment for neutropenia. Overall, only three patients in the observation arm and 10 patients in the RTX arm were recorded to have received colony stimulating factors for an adverse event during the maintenance/observation or follow-up phase. (See also Section 5.9.2.11.2 for laboratory assessments of neutrophil counts.)
Cardiac events	188	Cardiac AEs were recorded for six patients in the observation arm and 16 patients in the RTX arm. Of these, two patients in the observation arm and 11 patients in the RTX arm experienced cardiac disorders that were considered to be SAEs In addition, one patient (patient 10634/1036) experienced a cardiac event (arrhythmia) between randomization and the first observation visit and was therefore categorized as experiencing the event during the induction phase Two of the three SAEs in the observation arm were considered to be unrelated to trial treatment, whereas in the RTX arm three of four SAEs of cardiac failure (probable), one SAE of cardiomyopathy (possible), and one SAE of myocardial infarction (remote) were considered to be related to trial treatment (trial treatment could mean induction therapy or maintenance RTX). Importantly, including both arms of the study, all except one patient (who developed aortic valve disease) had received R-CHOP (ie, anthracycline-containing therapy) as their induction treatment. Most of the patients also had other risk factors for cardiac disease. Despite the seriousness of the conditions, almost all the patients in the RTX maintenance arm were able to continue with their RTX treatment, suggesting that RTX was not thought to be the cause or an exacerbating factor for their condition.
Laboratory parameters Hematology (including neutrophil counts) and biochemistry parameters were very similar between the two arms during the course of the maintenance/observation phase, with the exception of lymphocyte counts (Section 5.9.2.11.1).		
Lymphocyte counts	196	Lymphocyte counts increased with time in the observation arm compared with the RTX arm This difference was probably due to B-cell recovery in the observation arm compared with continued B-cell suppression in the RTX arm

Analysis	Page in MS	Manufacturer's summary as presented in the MS
Neutrophil counts	197	Neutrophil counts were very similar in both arms throughout the maintenance/ observation phase. In both arms, mean and 95% confidence intervals returned to the normal range by visit 2 and remained within this range thereafter.
Shifts from baseline	197	The majority of patients in both study arms showed no change in NCI-CTC grade for any laboratory test parameter during the maintenance/observation phase. The number of patients whose laboratory values worsened during the maintenance/observation phase and shifted to NCI-CTC Grade 3/4 is summarized in. In the RTX arm, a higher number of shifts to Grade 3/4 values was observed for lymphopenia as well as leukopenia and neutropenia. There were very few shifts to Grade 3/4 for blood chemistry parameters, and for these parameters there was little difference between the two study arms.
Marked laboratory test value abnormalities	198	The most common Grade 3 or 4 hematological laboratory abnormalities were neutropenia and lymphopenia,. Grade 3 or 4 neutropenia and lymphopenia were more frequent in the RTX arm (9% and 13%, respectively) than in the observation arm (4% and 6%, respectively). There were fewer reports of adverse events of neutropenia or lymphopenia than Grade 3 or 4 decreases in neutrophils or lymphocytes, respectively). In the RTX arm, 9% of patients recorded Grade 3 or 4 neutropenia based on laboratory counts, but only 4% of patients had adverse events of neutropenia reported. Similarly, 13% of patients in the RTX arm experienced Grade 3 or 4 lymphopenia based on laboratory data, but less than 1% of patients had adverse events of lymphopenia reported. These disparities are typical in oncology studies and reflect the fact that short episodes of neutropenia and/or lymphopenia often have no adverse consequences for the patient.
Differential lymphocyte counts	200	Patients at study sites in France underwent additional sampling for immunophenotyping of peripheral blood cells. Absolute levels of circulating B-cells (CD19-positive), T-cells (CD3-positive), and natural killer cells (CD16- or CD56-positive) were assessed before induction therapy, after induction therapy (baseline), and every six months for the first three years after randomization or until recovery if not reached at this time.
B-Cells	200	Analysis of CD19-positive lymphocyte subsets showed suppression of B-cells in both study arms at baseline (after completion of induction therapy) and continued B-cell suppression during the maintenance/observation phase for patients in the RTX arm. In comparison, patients in the observation arm showed recovery of B-cells during the maintenance/observation phase, with the mean value returning to within the normal range by visit 6 (ie, approximately one year after completing induction therapy). The mean B-cell count in the observation arm at the end of the maintenance/observation phase was 0.16×10^9/L.
T-Cells	201	The mean T-cell counts at baseline (after completion of induction therapy) in both study arms were similar and within the standard reference range (mean 0.90×10^9/L in the observation arm, and mean 0.91×10^9/L in the RTX arm). Although the mean values increased slightly in the observation arm and decreased slightly in the RTX arm at visit 3 (mean 1.06×10^9/L in the observation arm, and mean 0.86×10^9/L in the RTX arm), there was little difference between the two arms over subsequent visits and most patients in the two arms remained within the normal range throughout the maintenance/observation phase).
Natural killer cells	202	The mean counts of natural killer (NK) cells at baseline (after completion of induction therapy) in both study arms were similar (mean 0.17×10^9/L in the observation arm, and mean 0.17×10^9/L in the RTX arm) and increased slightly during the course of the maintenance/observation phase. At visit 12, mean NK cells counts were 0.24×10^9/L in the observation arm and 0.23×10^9/L in the RTX arm.
Serum immunoglobin levels	203	Patients at study sites in France also underwent additional sampling for immunoglobulins (IgG, IgA, and IgM). Immunoglobulin levels were assessed before induction therapy, after induction therapy (baseline), and every six months for the first three years after randomization or until recovery if not reached by this time.

Analysis	Page in MS	Manufacturer's summary as presented in the MS
Immunoglobin G	203	The mean IgG levels at baseline (after completion of induction therapy) in both study arms were within the reference range of 5.00–12.00 g/L (mean 7.76 g/L in the observation arm, and mean 7.87 g/L in the RTX arm). Over the course of the maintenance/observation phase, the mean values in both study arms remained within this reference range however, there was a slight decrease in mean IgG levels and 95% confidence intervals in the RTX arm over time compared with the observation arm, although there was still considerable overlap. At the end of the induction phase, IgG levels were very similar between the two arms): 113 of 126 patients (90%) in the observation arm and 116 of 131 patients in the RTX arm (89%) had IgG levels within the reference range (5.00–12.00 g/L). Forty-five patients (36%) in the observation arm and 46 patients (35%) in the RTX arm had IgG levels lower than 7 g/L. Ten patients (8%) in the observation arm and 11 patients (8%) in the RTX arm had IgG levels below the lower limit of normal (5 g/L) after induction; of these, six patients (5%) and three patients (2%), respectively, had IgG levels lower than 4 g/L. Although the numbers of patients with available IgG data decreased during the maintenance/observation phase, the majority of evaluable patients in both arms continued to have IgG levels of 4 g/L or higher. At the end of the maintenance/observation phase, 11 patients (26%) in the observation arm and 36 patients (47%) in the RTX arm had IgG levels lower than 7 g/L. Only one patient (2%) in the observation arm and four patients (5%) in the RTX arm had IgG levels lower than 4 g/L at the end of the maintenance/observation phase. Recovery of IgG levels during the maintenance/observation phase was observed for five of the 11 patients who had IgG levels at the end of induction lower than the lower limit of normal (LLN) and lower than the value at screening (1/5 patients [20%] in the observation arm, and 4/6 patients [67%] in the RTX arm)
Immunoglobin A	206	The mean IgA level at baseline (after completion of induction therapy) was 1.57 g/L in the observation arm and 1.33 g/L in the RTX arm (reference range: 0.5–3.5 g/L). The mean values remained slightly higher in the observation arm over time compared with the RTX arm. Overall, mean IgA levels and 95% confidence intervals in the RTX arm overlapped those in the observation arm throughout the maintenance/observation phase, and no major differences were observed between the two arms In both arms, mean IgA levels and 95% confidence intervals remained within the normal range throughout the maintenance/observation phase. Seven patients (four patients in the observation arm, and three patients in the RTX arm) had IgA levels at the end of induction that were lower than the LLN and lower than at screening. None of these patients had a recovery in their IgA levels during the maintenance/observation phase
Immunoglobin M	207	The mean IgM level at baseline (after completion of induction therapy) was 0.59 g/L in the observation arm and 0.64 g/L in the RTX arm (reference range: 0.30–2.30 g/L). The mean IgM values increased slightly in the observation arm and decreased slightly in the RTX arm during the course of the maintenance/observation phase. However, overall, mean IgM levels in the RTX arm overlapped those in the observation arm throughout the maintenance/observation phase and no major differences were apparent between both arms. More patients in the observation arm (9/20 patients [45%]) had recovery in IgM levels during the maintenance/observation phase compared with the RTX arm (2/18 patients [11%])

Appendix 5

Table 33 EORTC 20981 trial characteristics

Study Name	Study intervention(s), comparator(s), drug, dose(s) and follow-up	Study design, location and enrolment	Study inclusion criteria	Study exclusion criteria	Study outcomes
EORTC 20981[29, 30]	Induction: N=466 R-CHOP: (N=234) Cyclophosphamide 750 mg/m^2 IV Day 1, doxorubicin 50 mg/m^2 IV Day 1, vincristine 1.4 mg/m^2 (maximum 2 mg) IV Day 1; RTX 375 mg/m^2 as a slow IV infusion Day 1, prednisolone 100 mg orally, Days 1 to 5 every 21 days. CHOP (N=231): As R-CHOP without rituximab Maintenance treatment: N= 395 RTX (N=167) RTX 375 mg/m^2 as a slow IV infusion once every 3 months for 8 doses (24 months) or until disease progression. Observation (N=234)	Open label randomized phase III trial Patients randomized to induction with 6 cycles of CHOP+/-R. Patients achieving PR/CR after 6 cycles randomised to 2 years of maintenance with RTX or observation. International study (130 centres) including UK 37 UK study centres. Enrolment between: November 1998 and April 2004	**Induction phase** • Ann Arbor Stage III or IV fNHL • Relapsed disease after a maximum of one or two adequate non-anthracycline containing CTX regimens • No prior treatment with anthracyclines, mitoxantrone or RTX • Circulating tumour cells < 10 x 10^9/L • Remission to at least one of the prior chemotherapy regimens (modified in June 2000 to include patients with stable disease as their best prior response) • Response duration of 3 months or more to one prior CTX • CD20 positive fNHL • At least one bidimensionally measurable lesion • Age ≥18 years and written consent • WHO Performance status 0, 1, or 2 **Maintenance phase** • Complete or partial remission (CR or PR) of at least 4 weeks duration after the last cycle of CHOP+/-R • For patients receiving RTX during remission induction: no RTX-related toxicity necessitating stopping RTX. • 4-8 weeks since last cycle of CHOP+/-RTX • IgG levels <6g/L (reduced to 3g/L in June 2000) • No active infection.	**Induction phase** • Severe cardiac disease • Serum creatinine, BUN, alkaline phosphates or bilirubin =/> 2.5 times the upper limit of normal, unless clearly related to lymphoma • Pregnancy • Prior malignancy, except non-melanomatous skin cancers, cervical carcinoma *in situ* and cancers cured by surgical resection > 5 years ago. • HIV positivity • Uncontrolled asthma • IgG levels <6g/L • Prior stem cell transplantation • Planned peripheral blood stem cell collection using CTX for mobilisation. **Maintenance phase** • None stated	• Last tumour response rate (LEXCOR criteria) • Progression-free survival • Overall survival • Event-free survival • Time to new anti-lymphoma treatment or death • Disease-free survival • All AEs regardless of causality occurring during or up to 30 days after the last treatment cycle/observation period.

Appendix 6

Table 34 Quality assessment of EORTC 20981 trial

Criterion	ERG assessment
Concealment of allocation	Open-label study. Placebo control for a study involving IV rituximab administration and oral and IV pre-medication would been very difficult and probably considered unethical during maintenance/observation Endpoints are fairly objective and placebo effect not likely to be a major problem
Randomisation technique	An appropriate technique was used: centralised using minimisation approach of Pocock and Simon
Sample size justified adequately?	Yes. Though given there was no data on which to base any assumption on the efficacy of rituximab maintenance at the time of protocol development
Adequate follow-up	Yes . Closure of first- and second-randomisations was mandated by independent monitoring given highly statistically significant differences in outcomes, making further follow-up very unlikely to change outcomes materially. Follow up appropriate to trajectory of disease with follow-up being longer than the median PFS after induction therapy
Assessors aware of treatment allocation?	Unclear. Although no reference made to blinding of assessors it is likely that scan results which would determine response/progression would, in most cases, be reported by radiologists with no interest in the study
Parallel group/cross-over	Parallel-group. Primary endpoints in both parts of study not influenced by post-study treatment
Carried out in UK?	International study including UK. There were 37 UK study centres who recruited 102 of the 465 patients randomized. Indicating that UK clinicians found this study pertinent to their practice and had plenty of patients fitting the study entry criteria within their clinical population. The control treatment in this study (induction with CHOP followed by no further treatment until relapse) is used in routine clinical practice in the UK
Dosage regimen	For both induction and maintenance portions of the trial dosage regimens accord with SmPC recommendations
Study groups comparable?	Yes
Appropriate statistical tests?	Yes. Note that statistical analysis in this study has been subjected to both peer-review for publication and EMEA scrutiny
ITT analysis?	Yes for both induction and maintenance portions of study

Appendix 7

Table 35 Unit costs and adverse events costs used by the manufacturer

Model variable	Value	Source
Unit costs		NHS Ref Costs 08-09 (for codes used, see MS, Table 103)
Hospital clinic visit (haematologist)	£131	
Hospital clinical visit (junior)	£83	
CT scan	£157	
Full blood count	£3	
Patient history	£4	
Liver function test	£4	
Urea & electrolytes	£4	
Immunoglobulin	£8	
Bone marrow biopsy	£26	
Lactate dehydrogenate test	£4	
HIV serology	£3	
Hepatitis serology	£3	