[redacted]

7th January 2011

Jeremy Powell
MidCity Place
71 High Holborn
London
WC1V 6NA

BY E-MAIL

Re: Appraisal Consultation Document - Rituximab for the maintenance treatment of follicular non-Hodgkin's lymphoma following response to first-line chemotherapy

Dear Jeremy,

Thank for providing us with the Appraisal Consultation Document. Please find below Roche's response to the ACD. The results of the additional analyses requested by the Committee demonstrate that rituximab for the treatment of 1st line maintenance is a cost effective use of NHS resources under plausible scenarios tested. It is important to note here that the scenarios which resulted in am ICER of more than £30K per QALY are based, in Roche's opinion on implausible clinical assumptions and should be viewed with caution. In summary the ICER is above £30 per QALY if all of the criteria below are met:

1. Rituximab's **treatment effect lasts for only 36 months** – This is not consistent with **all** available trial data for rituximab across all licensed indications (NHL, CLL, DLBCL)
2. Clinical outcomes for patients that received 1st line maintenance are considerably worse than those observed in the EORTC trial upon relapse and subsequent treatment with 2nd line therapies. **This results from artificially changing model input parameters to generate the 50% PFS to OS conversion rate**
3. The age of the cohort at the start of treatment is 65 years old

If you require any further information or clarification then please do not hesitate to contact us.

Yours Sincerely,

[redacted]

A. Has all of the relevant evidence been taken into account?

The analyses requested by the Committee in the ACD has resulted in a wide range of ICERs. Roche attempted to address all concerns and remaining uncertainties of the cost effectiveness of rituximab in 1st line maintenance treatment of NHL. In the one-way sensitivity analysis **none** of the issues flagged by the committee was found to have a significant impact on the cost effectiveness.

The range of ICERs were as follows:

- 'Conversion rate': £15,978 – £19,339 per QALY
- Treatment effect duration: £15,978 – £26,079 per QALY
- Age of the cohort: £15,978 – £16,645 per QALY
- Increased utility due to delay of chemo: Positive impact on ICER

It is important for the Committee to understand in reasonable detail how Roche altered the model in order to run these sensitivity analyses and obtain the ICERs in the upper range of the spectrum. Model parameters had to be altered to very extreme values, contradicting the available clinical data. This was done with the sole purpose of artificially simulating the 'conversion rates' requested by the Committee, without evaluating the resulting clinical inputs/assumptions for their face validity. Roche maintain that the most plausible scenario is reflected in the base case analysis, as the assumptions on which it was developed were based on the best available scientific data.

B. Are the summaries of clinical and cost effectiveness reasonable interpretations of the evidence?

Clinical trial

On page 6 of the ACD, NICE refer to the "premature closure of the [PRIMA] study".

Roche would like to clarify that the study was not stopped early. The study was stopped because the independent DSMC (in Sept 2009) declared that the study had reached its primary endpoint at the pre-specified interim analysis.

It is also worthy to note that the statistical hypothesis and analysis plan of the PRIMA trial were declared appropriate by the rapporteurs for the study (Danish and Dutch Medicines agencies) in 2006. The study was also declared adequately designed to support a filing for MabThera maintenance therapy in previously untreated follicular lymphoma (FL) patients. On the basis of these positive data the EMA granted a licence extension for rituximab as maintenance therapy in first-line FL on the 25th Oct 2010.

Modification of PRIMA primary endpoint from EFS to PFS

In the committee meeting on 4th November Roche were asked why the primary endpoint in the PRIMA study was changed in August 2006 from EFS to PFS. Expanding on our previous answer, this modification was recommended since both European and US Health authorities consider that PFS benefit is the primary criteria for drug registration. This was one of several amends proposed by Roche and Genentech to the international sponsor of the study (GELA) to justify the opportunity

to use the PRIMA data to file for a rituximab maintenance licence in previously untreated follicular lymphoma patients.

Cost effectiveness analysis

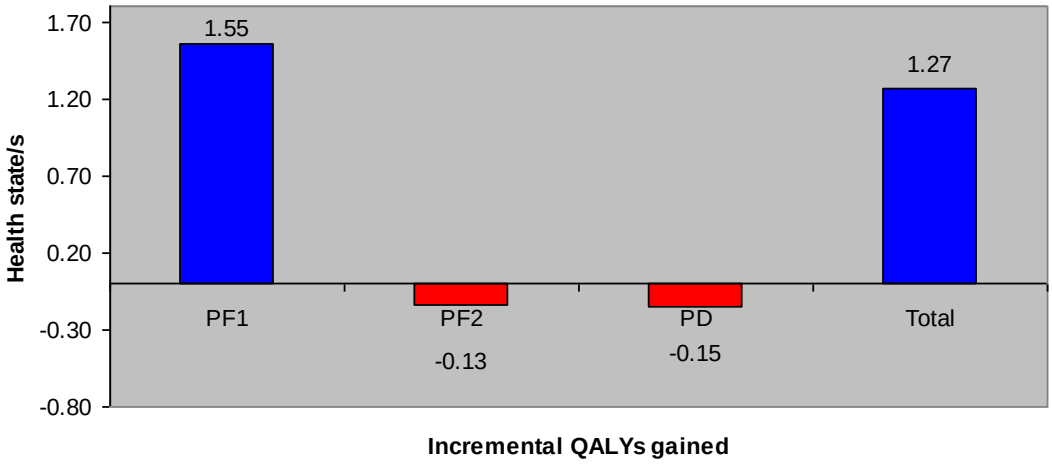
i) The amount of converted PFS to OS benefit is an artefact of the clinical inputs/assumptions

a) Explanation of the mechanics of the model

Roche would like to provide more information on the mechanics of the model in order to aid the Committee understand how the model 'converts' such a high percentage of the benefit gained in PFS into OS benefit. It is important to note here that the 'conversion rate' is not an input in the model (specific model parameter) but an output of the underlying modelling assumptions; additional QALYs gained in PF1 for the intervention arm (modelled from the PRIMA trial) and similar QALYs gained in both arms in the later stages of the disease.

The model assumptions with regards to progression from PF2 to PD, PF2 to death and PD to death are **identical** for the intervention and comparator arm. The only difference in clinical model inputs is the rate of progression from PF1 (1LM) to PF2. These progression rates were determined directly from the phase III PRIMA trial. This is the stage of the disease where rituximab in 1LM has its impact reflected in the PFS gain and the QALY benefit predicted by the model. No additional incremental benefit is accumulated in the late stages of the model for patients in the intervention arm. As the figure below demonstrates, the observation arm accumulates more benefit following the PF1 phase of the model.

Figure 1: Incremental QALYs gained in the 3 health states in the rituximab arm (base-case). The 4th column (right) represent the total incremental QALY gained



As it can be seen from the figure above, the conversion percentage have been incorrectly reported in the ERG report and section 4.8 of the ACD as being 96.6%. In fact the conversion rate of PFS QALY gained to total (OS) QALY gained is 81.7% (1.27/1.55 ×100%).

In order to artificially achieve a conversion rate that is less than the base-case predicted rate, the model has to be altered in a number of ways that may not reflect what has been observed in the EORTC 20981 trial (used to model later stages of the disease). It is important to recognise and agree on the plausibility of the underlying model inputs / assumptions utilised in order to achieve the lower conversion rates requested. The table below summarises the model parameters that drive the conversion rate and therefore could be subject to modifications in order to simulate the committee's requested scenarios.

Table 1: Overview of sensitivity analysis that could affect the conversion of PFS benefit to total benefit.

	PF2-PD progression		**PF2-death progression**		**PD-death**	
	R 1LM arm	Obs arm	R 1LM arm	Obs arm	R 1LM arm	Obs arm
Base-case	Same rate across 2 arms		Same rate across 2 arms		Same rate across 2 arms	
Potential scenario	Patients progress to PD faster	Patients progress slower	Patients die faster	Patients die slower	Patients die faster	Patients die slower

b) One-way sensitivity analysis utilising assumptions that lead to different PFS to OS 'conversion rates'

In an attempt to address the Committee's requests for extra analysis exploring a range of "conversion rates", 3 scenarios are presented. The target of each scenario is to reduce the treatment efficacy of subsequent lines for the rituximab arm only, in order to lower the conversion rate.

Scenario 1: Decrease the efficacy of 2 line treatments in the intervention arm by increasing the rate of progression from PF2 to PD.

Table 2: Sensitivity analysis altering the rate of progression from PF2 to PD for patients in the rituximab arm

Increase from base-case in the rate of progression from PF2 to PD	**0% (base-case)**	**10%**	**20%**	**30%**	**40%**	**46%¢**
Conversion rate	81.8%	73.1%	65.7%	59.0%	53.2%	50.0%
ICER (£ per QALY)	**15,978**	**16,632**	**17.326**	**18,0623**	**18,845**	**19,339**

Efficacy of 2nd line therapies worsening for intervention arm compared to comparator arm. Despite patients receiving the same treatment as in the observation arm patients need **to progress from PF2 to PD almost twice as fast** in order to achieve the conversion rate of 50%

Scenario 2: Decreasing the efficacy of 2 line treatments in the intervention arm by increasing the rate of death while in PF2 for patients that received rituximab 1LM.

Table 3: Sensitivity analysis altering the rate of dying from PF2 for patients in the rituximab arm

Increase from base-case in the rate of progression from PF2 to death	**0% (base-case)**	**100%**	**200%**	**300%**	**315%¤**
Conversion rate	81.8%	71.59	61.2%	51.4%	50.0%
ICER (£ per QALY)	**15,978**	**16,620**	**17,442**	**18,420**	**18,582**

Efficacy of 2nd line therapies worsening for intervention arm compared to comparator arm. Despite patients receiving the same treatment as in the observation arm **patients need to die while in PF2 more than 4 times as fast** in order to achieve the conversion rate of 50%

Scenario 3: Decreasing the efficacy of 2nd line treatments by increasing the rate of death in PD for patients in the intervention arm

Table 4: Sensitivity analysis altering the rate of dying in PD for patients in the rituximab

Increase from base-case in the rate of progression from PD to death	**0% (base-case)**	**10%**	**20%**	**30%**	**40%**	**46%§**
Conversion rate	81.8%	72.9%	65.3%	58.6%	52.8%	50.0%
ICER (£ per QALY)	**15,978**	**16,473**	**16,968**	**17,461**	**17,951**	**18,205**

Efficacy of 2nd and 3rd line therapies worsening for intervention arm compared to comparator arm. Despite patients receiving the same treatments as in the observation arm patients **need to die while in PD almost twice as fast** in order to achieve the conversion rate of 50%.

c) Summary of analyses by artificially altering assumptions to achieve a conversion rate of 50%

Roche was unable to estimate the ERG's base-case conversion rate nor replicate ERG's sensitivity analysis around the conversion rate. However a wide range of sensitivity analysis was presented that the ICER remains comfortably below the £20,000 per QALY threshold.

It is important for the Committee to recognise the assumptions required within the model in order to generate some of the most extreme scenarios. In order to achieve a conversion rate of 50.0% the efficacy of 2nd line treatments for patients previously exposed to rituximab maintenance has to be substantially different to what was observed in the EORTC 20981 trial. Most notably, in the examples derived from the sensitivity analysis conducted above:

To achieve a conversion rate of 50% patients that received rituximab in 1st line maintenance:

- Need to progress from in PF2 to PD at a rate that is **150%** faster than that observed in trial. (**scenario ¢ in table 2 above**)
- Need to die while in PF2 at a rate that is **~400%** faster rate than that observed in trial. (**scenario ¤ in table 3 above**)

- Need to die while in PD at a rate that is **~150%** faster rate than that observed in trial. (**scenario § in table 4 above**)

Roche questions the plausibility of such scenarios. The wealth of data have consistently shown that the above underlying assumptions are substantially worse compared to the what has been observed in the clinical practice and clinical trial setting. As presented in the original submission rituximab's 2nd line treatment effect in treatment experienced patients is maintained (see 6.3.1 of the submission; p253-254). Studies[1,2,3] demonstrated that patients experience a similar treatment benefit in PF1 and PF2 health states. The base-case model utilises the best available data to model coming from the phase III trial in 2nd line treatment to model PF2.

ii) Extend of treatment effect

a) Sensitivity analysis

Roche present below the findings of the one-way sensitivity analysis to address the Committee's concerns with regards to the duration of the treatment effect. The conversion of PF1 benefit gain to total benefit is also shown in the table, as this is another area of concern and to demonstrate the interaction between the 2 underlying assumptions.

Table 5: Treatment effect duration and impact on the ICER

Duration of rituximab 1LM treatment effect	**72 months (base-case)**	**36 months**	**48 months**
Conversion rate	81.8%	91.9%	86.7%
ICER (£ per QALY)	**15,978**	**26,079**	**20,841**

It is important to note here that the underlying assumption is that patients will stop exhibiting the effect of rituximab maintenance after the pre-specified period of time. After that period the hazard ratio is assumed equal to 1. This is inconsistent to what has been observed from all the available data for rituximab (table 6 below).

While the Committee judged the treatment effect during the Appraisal Committee meeting by 'eye-balling' the Kaplan-Meier curves, the cumulative hazard plot which provides a much clearer picture of the relative hazard rates was not evaluated during the meeting but was reported in the original Roche submission (Figure 32; p258). We present the graph (originally shown in the submission) showing that the cumulative hazard curves are not parallel therefore demonstrating that the treatment effect is maintained long after patients stop being treated with rituximab. Based on the

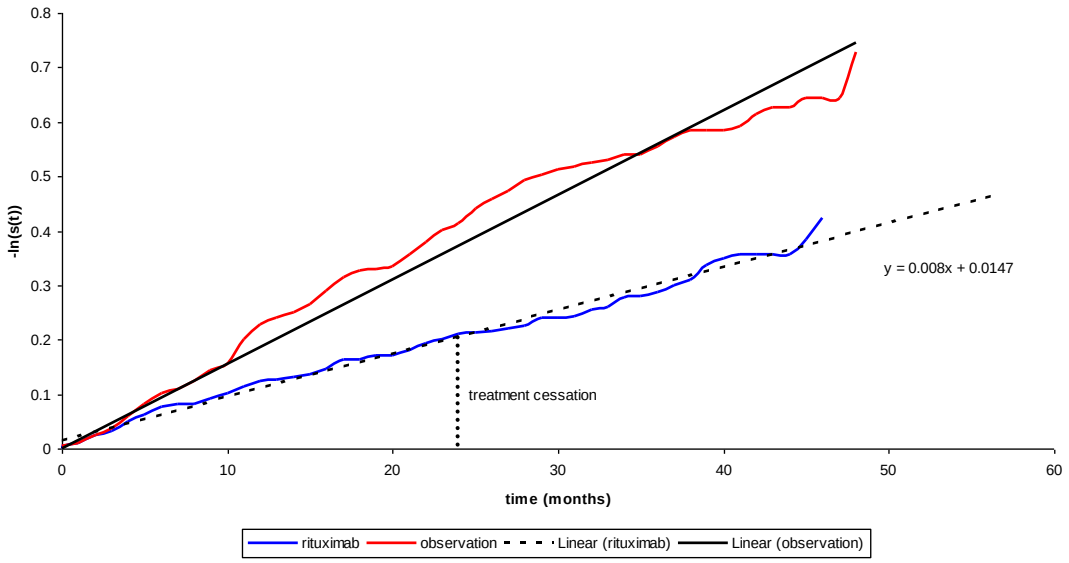
[1] Johnston, A., et al. (2010). Retreatment with rituximab in 178 patients with relapsed and refractory. *Leukemia & Lymphoma*. March 2010; 51(3): 399–405
B-cell lymphomas: a single institution case control study

[2] Coiffier, B., et al (2002). Rituximab re-treatment in B-cell lymphomapatietns: efficacy and toxicity in 59 patients treated in one centre. *Blood*. 100:a1390 (abstract)

[3] Cohen, Y., et al (2003). Re-treatment with rituximab alone induces sustained remission in a patient with follicular lymphoma with multiple extrnodal sites of involvement, relapsing soon after PRIMAry treatment with fludarabine-rituximab. *The Hematology Journal*. 4:151-3

evidence below it would seem unreasonable to assume the treatment effect ceases after 36 months.

Figure 2: Cumulative hazard plots for the 2 arms of the PRIMA study



The data shown in the PRIMA trial (above) and previous rituximab trials in FL, are in contrast to the Committee's suggested assumption for the treatment effect parameter of the model. Roche would like to highlight, data from other rituximab trials in FL with longer follow-up than PRIMA to demonstrate the treatment effect is maintained post treatment cessation with marginal increase on the observed hazard ratio. In the Marcus trial the treatment effect was observed for **4 years** post last scheduled treatment. Consistently, in the EORTC trial, rituximab's clinical benefit over time was shown to last for at least **4 years** post last scheduled dose. Given the consistency observed in previous trials and the cumulative hazard plot from the PRIMA study, applying the observed HR PRIMA for less than 2 years post treatment cessation (4 years from start of trial/economic model) would unfairly bias the long term efficacy assumptions against rituximab.

Table 6: Rituximab trial data showing that the treatment effect is maintained for a long period of time showing marginal differences between early and later cut-offs

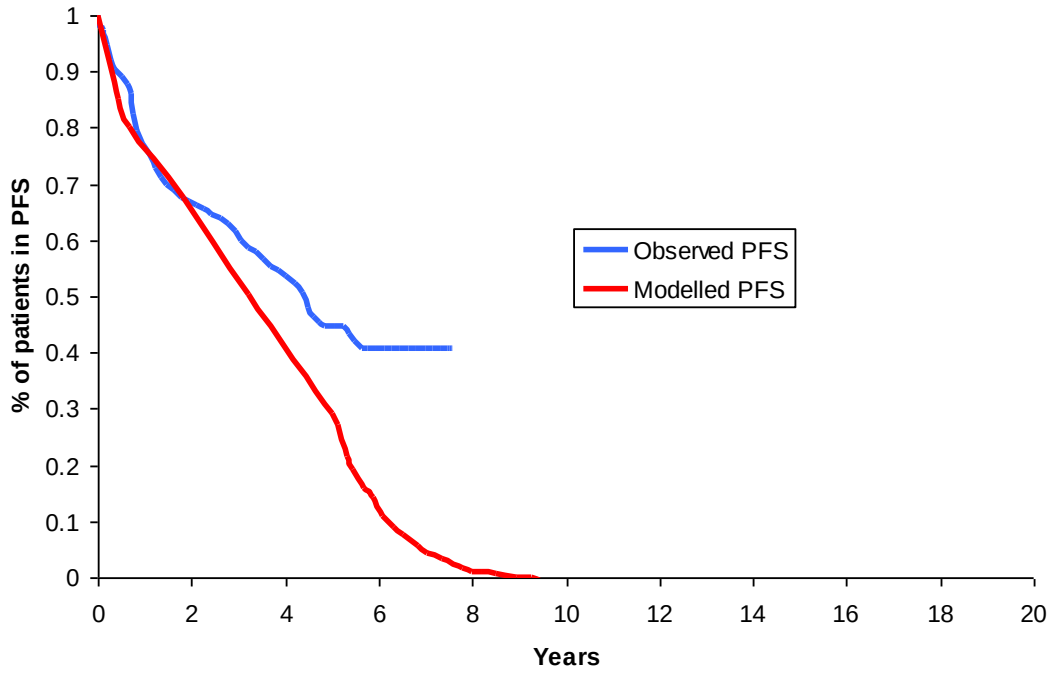
Study	Time from last scheduled rituximab dosing	HR	Length of follow up
Marcus R-CVP (n=322)	1 year	0.34	18 months
	2 years	0.40	30 months
	~4 years	0.44	53 months
EORTC 20981 (n=334)	1 year	0.40	33 months
	4 years	0.55	72 months
PRIMA (n=1018)	0 years	0.50	25 months
	2 years	0.55	47 months

It is also important to highlight that in the long-term follow up data from the EORTC studies has confirmed the outcomes of the extrapolated models included within Roche's previous submission (TA 137). In actual fact, observed data from long-term follow-up in this study suggests that modelled outcomes in the original submission

may even be *underestimating* the treatment effect using similar modelling techniques to those used in the first-line maintenance submission (figure below).

In the figure below the red line represents the modelled PFS as presented in the original submission for TA 137. The red line represents the PFS KM curve as observed in the latest follow-up of the EORTC 20981 trial (June 2010). The overlaying of the 2 curves (originally predicted and actual) clearly shows that the original model may have underestimated the benefit that rituximab actually was shown to deliver.

Figure 3: Modelled R-CHOP-R PFS (predicted in 2007 submission) vs actual PFS (June 2010) – EORTC 20981



A similar validation exercise was performed on the OS data of the EORTC trial, presented in appendix 1. The results show that the modelling techniques employed predict future trial results closely.

iii) Age

a) Sensitivity analysis by changing the age at the start of the model

In attempting to address the Committee's request for an analysis that is based on a representative for the UK age of the patient population, it is important to highlight the best methodology for presenting such analysis.

In the original Roche submission, an analysis of the treatment effect for first-line rituximab maintenance in 2 age groups, <60 years and >=60 years, was provided. This analysis was predefined in the PRIMA study protocol and patients were stratified by this baseline characteristic. The forest plot illustrating the hazard ratios for PFS with 95% confidence intervals (observation vs rituximab) for pre-specified patient subgroups is shown in Figure 9 of our submission (also found in appendix 1 of the ACD response). Results show that the risk of disease progression or death was

significantly reduced in the rituximab arm compared to the observation arm irrespective of age (<60 HR 0.45 95% CI 0.33-0.62; ≥60 HR 0.59 95% CI 0.39-0.90). This pre-specified analysis clearly demonstrates that the treatment effect does not differ between the 2 groups.

Any other, non pre-specified sub-group analysis of the trial data based on the age of the patients at the start of the trial would break randomization and the resulting findings may be confounded by other imbalances between patient baseline characteristics.

In order to provide an insight on the sensitivity of the economic model to the age parameter the cost effectiveness analysis was performed by altering the age of the cohort in at the start of the model, thus altering the all-cause mortality rate. The ITT HR and base-case extrapolations of PFS were used for all the model re-runs.

Table 7: Sensitivity analysis demonstrating the cost effectiveness of rituximab 1LM compared to observation. The various scenarios examine the impact to the cost effectiveness when the age of cohort at the start of the model is increased

Age of the population at the start of the model	56 years (base-case)	58 years	60 years	62 years	64 years	65 years
Conversion rate	81.8%	80.1%	78.0%	75.5%	72.5%	70.9%
ICER (£ per QALY)	**15,978**	**16,073**	**16,197**	**16,350**	**16,537**	**16,645**

Clearly changing the age of the cohort at the start of the model has a limited effect on the cost effectiveness. This was expected given that the only parameter that is affected is the background mortality rate of the cohort.

iv) Utility increase associated with delaying relapse

Roche acknowledge that the model developed for the purposes of this submission does not explicitly capture the utility increase associated with the delay of subsequent chemotherapy treatment at relapse. It is evident that rituximab in 1st line maintenance delays relapse and therefore subsequent chemotherapy treatment, thus the current base-case ICER could be overestimated.

In attempting to address the short-coming of the model with respect to this parameter, several methodologies were considered for implementation in the current structure of the model. This was also discussed with the NICE technical team and ERG.

The model utilised is a cohort Markov model (as per the majority of oncology models used in HTA) and does not track individual patients through the cycles of the model. As a result of this structure there is no functionality to 'know' when individual patients have transitioned from PF1 to PF2. The same applies for patients transitioning from PF2 to PD. For this reason methods that rely on tracking of patients could not implemented on the model. Several options were considered:

- **Altering the structure of model in order to track patients** – Changing the structure of the model may have offered more accuracy in terms of patients' quality of life throughout the different stages of the disease. To implement these type of changes would require a complete rebuild of the economic model. Given

the time constraints for response to this ACD the model could not be change to accommodate this changes.

- **Applying a utility decrement associated with 2nd line chemotherapy treatment** – Costs and benefits of subsequent lines of therapy are applied in the 1st cycle of the model. Any disutility associated with subsequent lines of therapy would apply in both arms of the model, applied in the 1st cycle and therefore would be balanced between the 2 arms. The cost effectiveness results would therefore be unchanged

- **Increasing the utility value associated with PF1 for patients that stay in this state for more than a specified period of time** – It was deemed that this method would reward the treatment arm as a greater proportion of patients remain in this state for longer period compared to the observation arm. Although this option increases the utility for patients that remain healthy in PF1 it was deemed not scientifically correct and not appropriate for use in HTA. The reason is that the proportion of patients that stay healthy for a longer period of time will experience an ever improving quality of life. Roche believe that this assumption is no plausible at face value.

Given the time constraints and limitations of the model Roche were unable to alter the model in a robust way and provide a sensitivity analysis to address the Committees requests for an analysis that incorporates a utility increase for delaying relapse and subsequent lines of treatments with chemotherapy. The methods outlined above could be implemented but would compromise the scientific integrity of the modelling approach. Roche though would like to acknowledge here that if this assumption was changed to reflect what was requested by the Committee it would have a positive impact on the base case ICER.

v) Cumulative scenarios

Roche provide here scenarios in which the assumptions analysed above are altered at the same time in order to show the cumulative impact on the cost effectiveness of rituximab 1LM. Three sets of analyses were conducted (table 8, 9, and 10). Each of these sets how the age, treatment effect and transition probabilities of PF2 to PD[4] would have to altered in order to achieve the specified 'conversion rate' of 50%, 75% and 100%.

Table 8: Conversion rate target 50%

	Base case	Cumulative sensitivity analysis scenarios			
Transition probabilities (% worse than in base case in intervention arm)	1	128%	130%	132%	135%
Age	56	65	60	65	60
Tx effect	72	36	36	48	48
Conversion rate	**81.8%**	**50.0%**	**50.0%**	**50.0%**	**50.0%**

[4] Changing the transition probability of PF2 to death and PD to death shows similar results.

ICER (£ per QALY)	**15,978**	**36,386**	**36,797**	**27,456**	**27,722**

Table 9: Conversion rate target 75%

	Base case	Cumulative sensitivity analysis scenarios			
Transition probabilities (% worse than in base case in intervention arm)	100%	104%	109%	101%	107%
Age	56	65	60	65	60
Tx effect	72	36	36	48	48
Conversion rate	**81.8%**	**75.0%**	**75.0%**	**75.0%**	**75.0%**
ICER (£ per QALY)	**15,978**	**28,310**	**28,593**	**22,125**	**22,313**

Table 10: Conversion rate target 100%

	Base case	Cumulative sensitivity analysis scenario
Transition probabilities (% better than in base case in intervention arm)	0%	13%
Age	56	56
Tx effect	72	72
Conversion rate	**81.8%**	**100.0%**
ICER (£ per QALY)	**15,978**	**14,929**

C. Are the provisional recommendations sound and a suitable basis for guidance to the NHS?

Following the Committee's comments and request for additional analysis, Roche has provided a wide range of sensitivity analyses in order to address all the remaining uncertainties in rituximab's cost effectiveness. All but a few extreme analyses resulted in an ICER of below £30K per QALY demonstrating that rituximab is a cost effective use of NHS resources in this setting. The scenarios in which the resulting cost effectiveness ratio was found to be above £30K per QALY rely on a set of assumptions that contradict what has been observed in well designed and controlled clinical trials for rituximab. As a result it is inappropriate for these extreme scenarios to inform the decision making process.

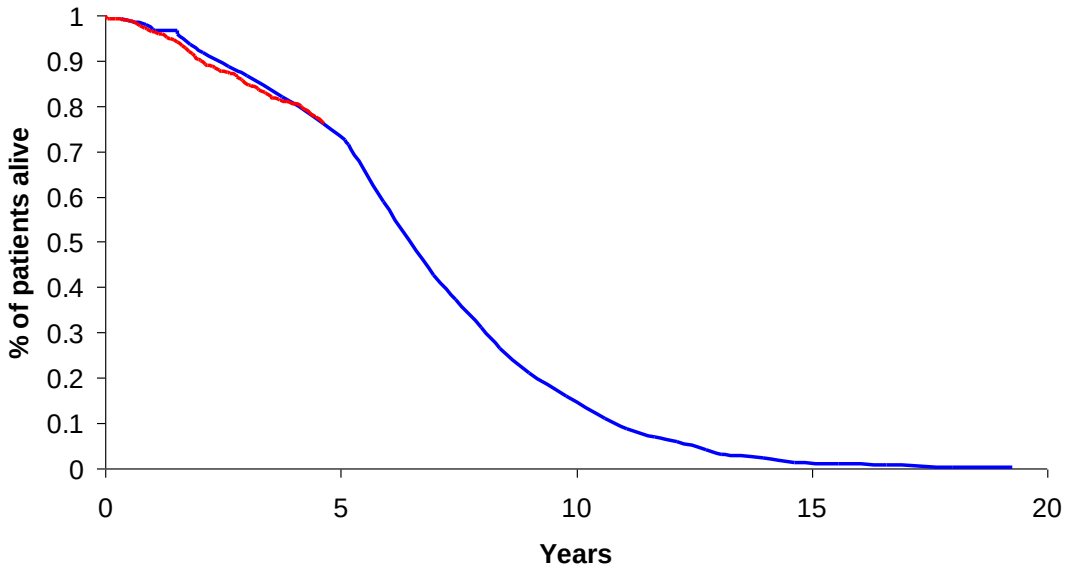
D. Are there any aspects of the recommendations that need particular consideration to ensure we avoid unlawful discrimination against any group of people on the grounds of gender, race, disability, age, sexual orientation, religion or belief?

None

Appendix 1

Roche conducted the same validation exercise as in figure 3 for OS from the EORTC trial by comparing the originally modelled curve for the R-maintenance arm (blue line) to the KM curves, as observed in the original and subsequent longer follow-up trial data. The figure below shows the original modelled OS benefit (based on the 33 month data). The red curve illustrates how the observed OS KM curve (red curve) from the 72 median follow-up data (van Oers 2010) subsequently tracked the original extrapolated curve very accurately.

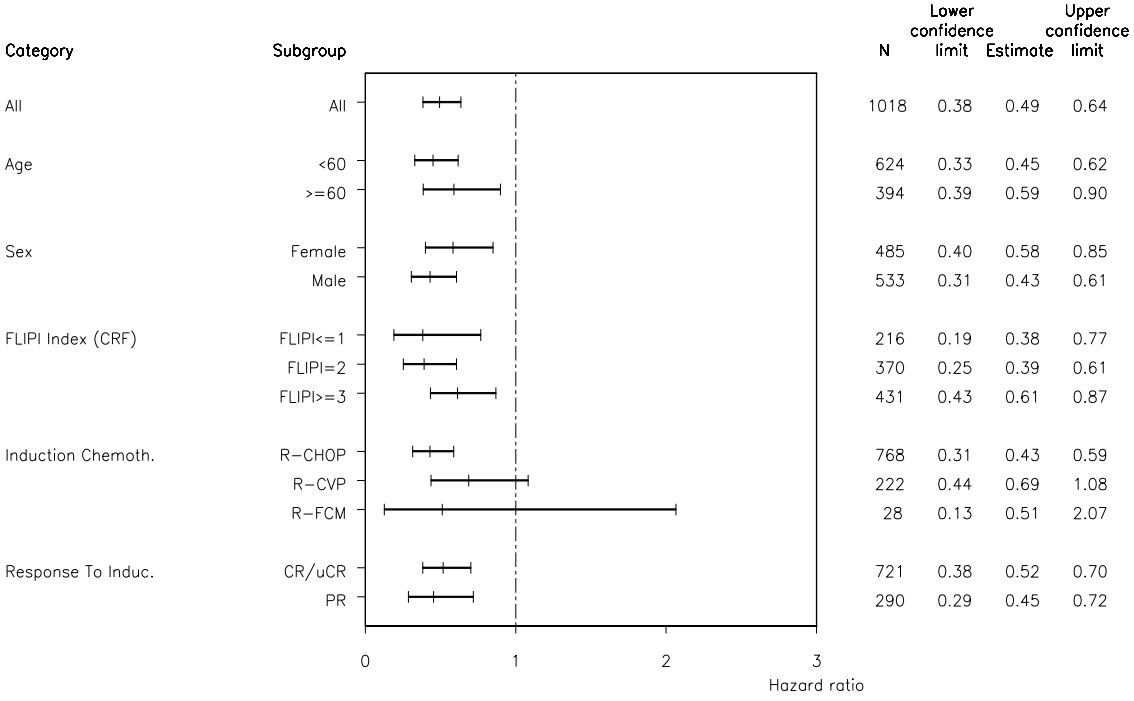
Figure 4: Modelled R-maintenance OS (predicted in 2007 submission) vs actual R-maintenance OS KM curve(June 2010) – EORTC 20981



Appendix 2

Figure 9 (original submission): Subgroup Analysis of Investigator-Assessed PFS (MITT)

Protocol(s): MO18264 (A18264M)
Analysis Population: MITT (N=1018)
Snapshot Date: 27OCT2009 Cutoff Date: 14JAN2009



PFS - day of randomization until 1st documented disease progression, relapse after response or death from any cause - investigator assessment.
Censoring occurs at last response assessment.

Program : $PROD/cd10752c/a18264a/mg_pfscox_hr.sas / Output : $PROD/cd10752c/a18264m/reports/mg_pfscox_hr_I.cgm
29OCT2009 9:11

After that period the hazard ratio is assumed to equal to 1 which is inconsistent to data observed