Appendix 2. Detailed cost-effectiveness response

As stated in the ACD document the ERG has noted that the core of the model appeared to be largely sound. However, the ERG had raised a number of concerns around the plausibility of some of the assumptions and inputs used in the economic model and also the Lilly approach to the modelling of progression free survival (PFS) and overall survival (OS). Here below we list all the key points raised as classified within the ACD:

1. Key changes to the basecase assumptions of the model:
    a. Increased frequency of monitoring (Section 4.10)
    b. Half-cycle correction for pemetrexed costs (Section 4.11)
    c. Cost of post-progression chemotherapy (Section 4.12)

2. Additional 6 minor alterations (Section 4.13)
    a. Cost assumptions
        i. Method for estimating pemetrexed costs using UK BSA data
        ii. Method for estimating docetaxel costs using UK BSA data
        iii. Concomitant medication costs (vitamin supplementation and dexamethasone)
        iv. Terminal care costs using a more up to date estimate from a recent report of first-line chemotherapy for NSCLC
        v. Inclusion of blood product cost

    b. Methodological assumption: unadjusted vs adjusted utility model

3. Methodological/Structural assumptions in the model
    a. The use of unadjusted versus adjusted OS and PFS models (Section 4.14)
    b. The cut off point at which a % of patients are at risk of death (Section 4.15)
    c. Approach to survival modelling: choice of parametric distribution and post-progression survival (PPS) gain (Section 4.16)

As stated in the cover letter we have incorporated many of the ERG suggestions to the model inputs. Please refer to the Table 1 below for an overview of the changes that have been incorporated. As can be seen from the above table when all those the changes that we considered plausible are incorporated the new basecase ICER is £58,918.

Of the *remaining 6 alterations* in the exploratory analysis of the ERG we have accepted all the changes suggested to the costs assumptions with the exception of the cost of vitamins and dexamethasone. Lilly believes that those should not be included as part of the basecase to maintain consistency with previous appraisals on pemetrexed (TA181 and TA190) where those were considered as part of the NHS Reference Cost for delivering chemotherapy. In addition, during 2012, Lilly introduced a free 'supplementation pack' service to hospitals in the UK (including vitamins, dexamethasone and patient information leaflets). The packs can be ordered by any hospital ordering pemetrexed at no additional charge. For both these reasons we do not agree they should be included in the revised basecase.

Of the methodological/structural ERG suggestions, we have also incorporated the amendment of the cut-off point at which a certain percentage of patients in both arms are assumed to be dead (25% versus 20% in the original basecase).

We have not, however, incorporated the remaining methodological changes in the basecase model as we do not agree that the methodology suggested provides a best approach in the economic evaluation. In order of their magnitude of effect upon the basecase, these are:

1) approach to survival modelling
2) Use of adjusted versus unadjusted for OS and PFS and
3) Use of adjusted versus unadjusted utilities

We explain our rationale for this below. In order to explain our rationale in a way that the Committee can understand the technical differences in the approach, a significant amount of detail is provided below.

Table 1 Model changes and results (based on the Lilly submitted model)

Index	**Original basecase analysis**		**£47,576**	
		Assumption incorporated	**Revised basecase**	**Change in ICER**
	Key changes to the model (ACD sections 4.10-4.12)			
1.a	Increased frequency of monitoring	✓	£49,126	£1,550
1.b	Half-cycle correction	✓	£50,552	£2,976
1.c	Cost of post-progression chemotherapy	✓	£51,332	£3,756
	Cumulative impact for accepted changes in Section 1		£55,948	£8,372
	Additional 6 minor alterations (ACD section 4.13)			
2.a.i	Method for estimating pemetrexed costs	✓	£48,854	£1,278
2.a.ii	Method for estimating docetaxel costs	✓	£47,707	£131
2.a.iii	Concomitant medication costs	✗	Not incorporated	£96
2.a.iv	Terminal care costs	✓	£47,518	£-58
2.a.v	Inclusion of blood product cost	✓	£47,576	No change to ICER
2.b	Unadjusted vs adjusted utility model	✗	Not incorporated	£1,659
	Cumulative impact for accepted changes in Sections 1 and 2		£57,251	£9,675
	Methodological/Structural assumptions in the model (ACD sections 4.14-4.16)			
3.a	Use of adjusted vs unadjusted parameters for OS and PFS	✗	Not incorporated	£1,991
3.b*	25% survival cut-point vs 20% at risk of death point (as per original submission)	✓	£48,953	£1,377
3.c	Approach to survival modelling	✗	Not incorporated	£13,936
	Cumulative impact for accepted changes, Sections 1, 2 and 3		£58,918	£11,342

**3.b uses 25% survival threshold to incorporate ERG's suggestion of equal survival by arm to avoid bias. A 25% cut-point means 42 patients (23%) in the placebo arm would remain at risk and 44 patients (12%) in the pemetrexed arm would remain at risk.*

3.c. Approach to survival modelling—choice of parametric distribution and post-progression survival gain

As stated in the economic section of our covering letter, the uncertainty regarding the assumption of OS treatment effect in the extrapolation period is predominantly a question of clinical plausibility. Here we focus on the modelling choices which have to be made in regards to the most appropriate parametric distribution and the assumption for post-trial treatment effect that impact PPS benefit.

Choice of parametric distribution

The ERG suggests that in our original submission we have overestimated the OS treatment effect of pemetrexed by using:

a) a misspecification of the survival function (i.e. gamma versus exponential) and
b) a single parametric function to generate OS estimates for both arms simultaneously.

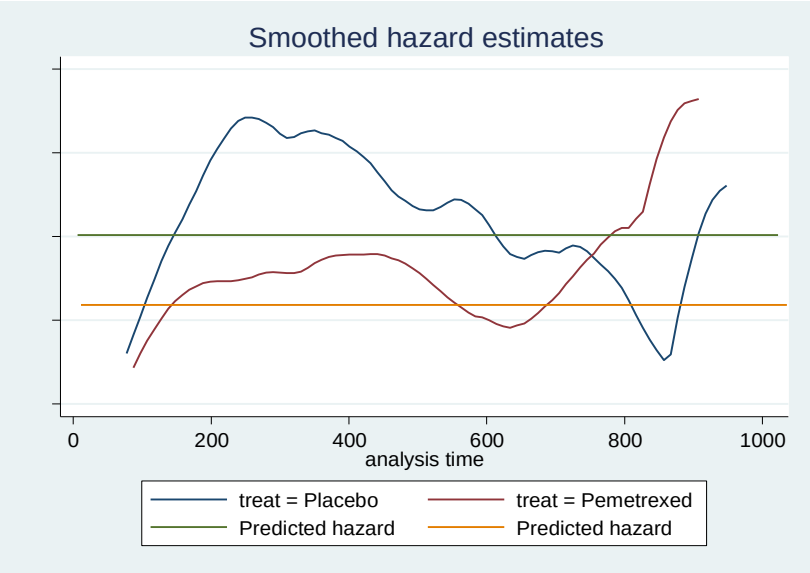
On the basis of our rationale below on both counts we believe that, by all conventional assessment the gamma distribution offers the best fit of the observed data compared to the exponential model and provides a clinically plausible long-term survival estimate. Hence, we stand by our original approach and have not incorporated the suggested changes on the model.

a) With regards to the *misspecification of the survival function* we believe the original choice of gamma is appropriate for the following reasons:
   - It better reflects the observed data in the PARAMOUNT Clinical trial. Evidence to support this argument is explained below by showing the smoothed hazard plots.
   - Standard statistical ways of assessing the goodness of fit also support the choice of gamma over exponential.
   - Residual plots provided by the ERG (Figures 8 and 9, ERG Report, p59, Jan 2013) show that the gamma distribution is a better fit over the whole trial duration and in particular towards the end of the trial where the extrapolation begins.

Smooth hazard OS estimates

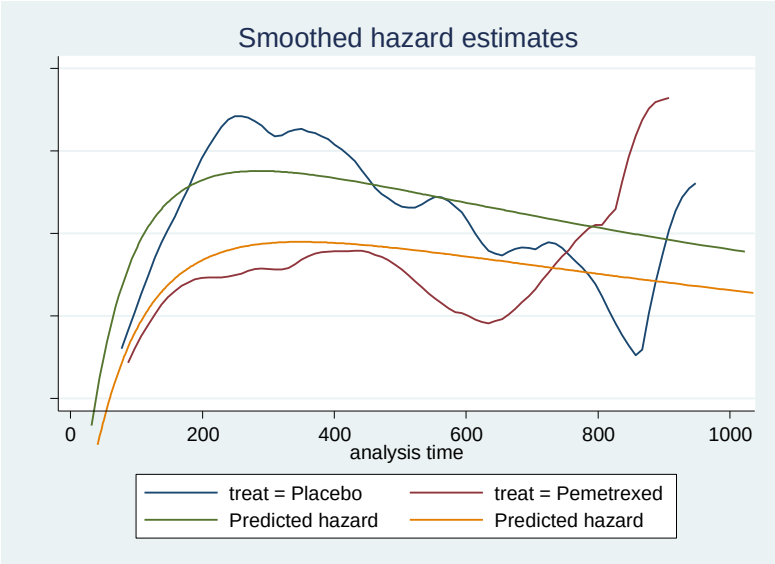
The parametric model used in the original submission took into account the underlying properties of the hazard rates in each treatment group, i.e. the instantaneous risk of death at different time points as recommended within the literature (Latimer, 2011). **An exponential function, as recommended by the ERG is depicted by the two horizontal lines in the Figure below and assumes that the hazard rate is constant over time.** The plot of the smoothed hazard rates for PARAMOUNT OS data (Figure 1) shows that in both treatment groups the hazard rate initially increases and then decreases over time i.e. it's not constant. An exponential model therefore fails to capture the initial increase and subsequent decrease in the baseline hazard evident in PARAMOUNT data and instead estimates a crude average hazard rate over the entire time period. This appears to be a less accurate or "good fit" to the observed survival data.

Figure 1 **Exponential model predicted hazard versus smoothed hazard OS estimates by treatment group over time**



The original model used the gamma model which is an Accelerated failure time (AFT) model, which predicts the underlying hazard to initially increase and then decrease, and therefore better reflects PARAMOUNT OS data compared to PH models such as exponential, see Figure 2. AFT models are able to model an underlying non-monotonic hazard, i.e. one which initially increases and then decreases, as is often seen in oncology trials. This appears a better fit to the observed PARAMOUNT OS data.

Figure 2 **Gamma model predicted hazard versus smoothed hazard OS estimates by treatment group over time**



Standard statistical analyses to test for goodness of fit

Other statistical analyses, presented in Lilly's original submission, indicate that the gamma distribution also offers a better fit of the data compared to the exponential distribution. The gamma model offers:

- lower AIC and BIC statistics; a lower value indicates a better model fit, (see Table 9 in Appendix 1.18)
- a better visual fit of the observed data as shown in plots of predicted survival curves against PARAMOUNT KM data (see Figures 12 and 17 in Appendix 1.18)
- a better fit based on Cox-Snell residuals, which compare the fit of each parametric model against a 45 degree line; estimates which lie closer to the line provide a better fit of the data (see Figures 11 in Appendix 1.18).

b) With regards to the use of a *single parametric function to generate OS estimates for both arms simultaneously*, the ERG has suggested that the approach used in the original submission was not appropriate because it may have introduced bias into the overall survival trendlines. It was postulated that using long-term OS exponential trends fitted independently to each treatment arm was a better approach.

The ERG suggestion is based upon a belief that trends observed in the latter part of the trial are suggestive of a constant hazard assumption. However, this suggestion is based upon estimates from a limited number of patients and events, data which are likely to be unreliable and introduce uncertainty.

Whilst we agree that it is necessary to establish appropriate survival projection, we believe that this projection should not be focused upon data which are likely to be subject to the greatest level of uncertainty at the end of the trial and that a projection using all of the observed data from the clinical trial (as per the original submission), is likely to be more reliable.

We have conducted additional statistical tests, which reinforce that the original choice of a gamma model which *used a single parametric function for both arms in the model* is appropriate for the following reasons:

a) It does not appear that the PH assumption is violated and as such it is not necessary to develop separate parametric models for each treatment arm.
b) A quantile-quantile (Q-Q) plot supports the use of a single AFT model to parameterise the PARAMOUNT OS data using the same approach for both arms

A PH model makes the assumption that *the treatment effect* is constant (proportional) to the baseline hazard. There are a number of tests that can be used to assess whether the PH assumption is upheld and we have provided results below on all these tests. These include:

- a test of the PH assumption (correlation between the Schoenfeld residuals and the rank order of failure events suggests PH violation) and inclusion of a time varying covariable in the regression model (evidence of a significant change in the hazard ratio over time suggests PH violation). (Table 2)
- a review of the hazard rates over time in each treatment group; (Figure 3)
- log-log plots, which map –log(survival) curves for each treatment against ln(time), (the PH assumption is assumed to be upheld if the curves are parallel); (Figure 4)
- a plot of the Schoenfeld residuals against time; ; (Figure 5)

The results of these statistical tests indicate that the **PH assumption is not violated** (p=0.74, Table 2 below) and the hazard ratio does not appear to vary over time (see figure 3).

Table 2 Test of PH violation

Test of PH assumption	rho	chi2	df	prob>chi2
Treatment	0.01619	0.1	1	0.7474
Global test		0.1	1	0.7474

Figure 3 Hazard ratio over time (estimated using a Cox model with a time varying covariable)

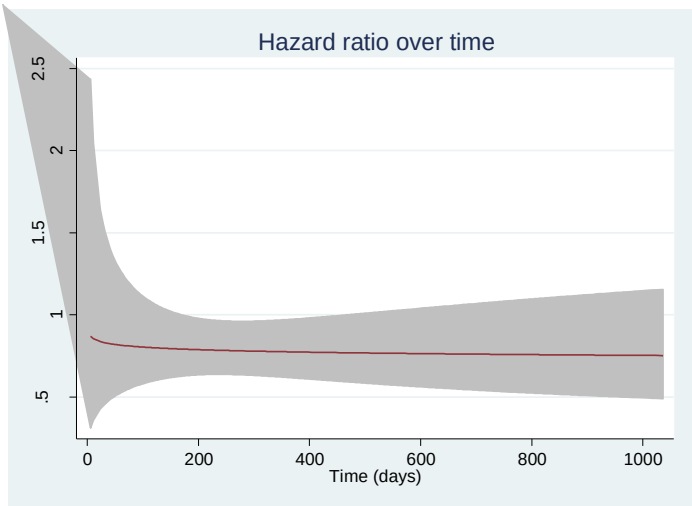


Figure 4 Log-log plots [–Ln(survival) Ln(time days)]by treatment

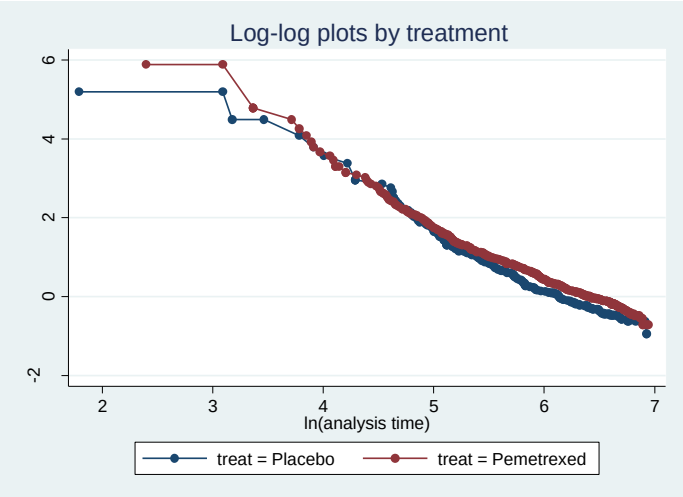
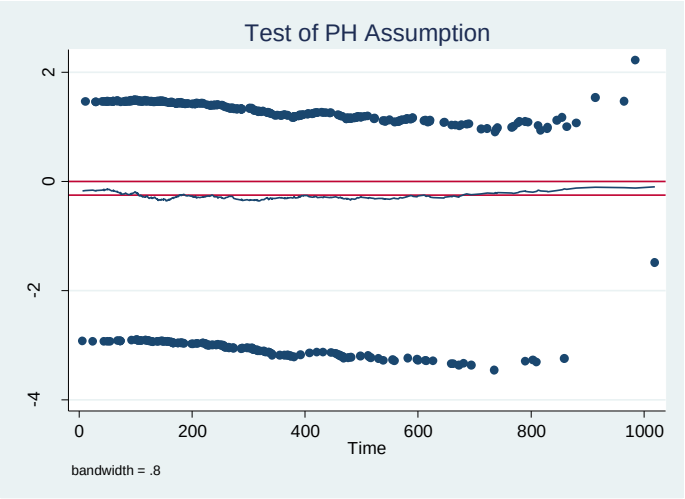


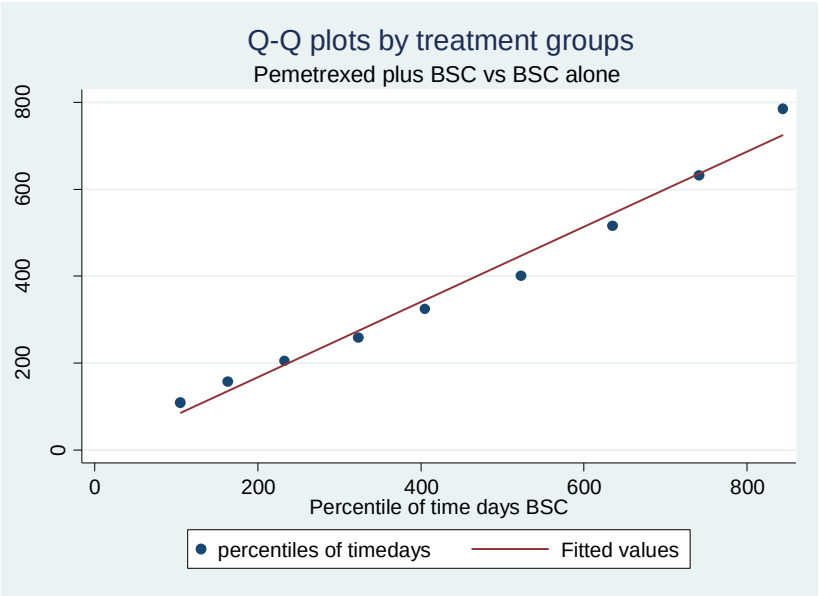
Figure 5 **Plot of the Schoenfeld residuals by treatment group**



It should be noted that the gamma model does not assume that the treatment effect is constant, i.e. proportional, to the baseline hazard. The gamma model, in fact, assumes that the treatment effect is multiplicative with respect to the underlying survival time and this functional form, together with the smoothed hazard plots, appears to appropriately characterise the relationship between the treatment arms in PARAMOUNT.

The best way to evaluate whether or not a treatment has a multiplicative effect on survival time (i.e. whether the AFT assumption is valid) is through quantile-quantile plots (Q-Q plots). In this instance Q-Q plots compare quantiles of the survival time for pemetrexed patients against quantiles of survival time for placebo patients. When the treatment effect is consistent across time points on this plot should be on an approximately straight 45 degree line (Patel 2006). Q-Q plots by treatment suggest that the AFT assumption is supported for PARAMOUNT data, see Figure 6.

Figure 6 **Q-Q plots by treatment groups**



Overall we believe that developing separate parametric models for each treatment group is not necessary and using a single-parameterised approach with an AFT model offers an appropriate solution. Furthermore, our supplementary statistical tests indicate that the AFT parameterisation is supported by PARAMOUNT overall survival data.

Post-progression survival

With regards to the *post-progression survival*, the ERG questioned the overall survival projection estimates for pemetrexed after disease progression and argued that no additional OS benefit should be accrued in the post-progression period. The ERG argued that the post-progression survival data from PARAMOUNT showed that the prognosis for patients in the post progression phase of both arms is the same (Figure 6, ERG report p57, Jan 2013). The ERG also argues that one of the main reasons for this bias in the OS benefit projection lies in the use that the gamma distribution. The ERG then argues that a revised exponential model would be a closer fit to the Kaplan Meier data.

From a pure modelling perspective, it is important to note that the model does not directly estimate post-progression survival, nor is post-progression survival (PPS) data used as a direct input into the model. Rather, PFS and OS estimates are based upon observed PFS and OS KM data for the within-trial period and extrapolated estimates in the post-trial period.

As stated in the cover letter of this response, the OS treatment effect in the model is based upon the within-trial period, taking account of both the pre-progression and post-progression periods. Uncertainty arises in relation to the treatment effect only in the extrapolation period and we therefore have to make an assumption in the model to this effect. By definition, no one knows with certainty what happens to the treatment effect after the trial ends. This is a question of clinical plausibility.

During the development of the model, (and supported during the Committee meeting), clinical experts suggested that it was biologically plausible that pemetrexed could alter the tumour such that treatment benefit continues post-progression. Therefore this benefit was included within the model. Whilst we do not have data specifically for pemetrexed, or for lung cancer, there is an emerging body of evidence that supports a hypothesis of biological plausibility for a continuing treatment effect post-trial for other cancers. It has been shown in advanced prostate and renal cell cancers (Stein et al 2009; Stein et al 2011) that part of the treatment effect is due to a change in growth rate of the residual tumour after treatment is stopped, i.e. the tumour has been fundamentally altered during treatment so that the tumour grows more slowly and survival is prolonged after the treatment is discontinued. The scenario proposed by the ERG that patients in the pemetrexed arm would, after disease progression, instantaneously get any treatment effect removed (i.e. treatment effect for pemetrexed patients in the extrapolation phase will be set equal to that of placebo) seems an extreme case and an unlikely clinical scenario given the biological plausibility of a continuing effect. It is therefore a conservative assumption representing the maximum ICER rather than the expected ICER.

We therefore believe that continued OS treatment effect in the extrapolation period is plausible and that modelling no beneficial effect is over conservative and likely to be inaccurate.

3.a The use of case-mix adjusted OS and PFS models

Lilly's original submission made full use of the Kaplan Meier data during the within trial period and then used an unadjusted model (which adjusts for treatment group only) to extrapolate the KM data beyond the trial duration. The ERG criticised the use of the unadjusted model and suggested that an adjusted model (which also adjusts for patients baseline characteristics) should be used to extrapolate PFS and OS because:

- The non-zero coefficient for some of the covariates in the adjusted regression model imply that some degree of bias is present in the estimated function (ERG report p54)

- The adjusted PFS and OS models are superior to the unadjusted models, explaining significantly more of the inter-patient variation and at least partially correcting for this modelling bias (ERG report p54)

On the basis of the above, the ERG then presented an alternative scenario analysis consisting of using the observed Kaplan Meier estimates from the PARAMOUNT trial up until the ERG's selected cut off point (i.e. 20% threshold common survival rate between both arms) at which the majority of the patients in the trial would already be dead by and subsequently appending an adjusted model to it to the end of the time horizon. This means that the extrapolated survival benefit used by the ERG uses a combination of both unadjusted model up to a certain point followed by an adjusted one for the remaining events.

An adjusted model may be preferred to an unadjusted model for consideration of subgroups in a cost-effectiveness analysis. However, no subgroups were identified during this appraisal and there was no clinical expectation that the treatment effect for pemetrexed would alter according to possible subgroup populations.

The ERG also states that the covariates used in the adjusted model exhibited statistically significant values indicating statistically superior model fit compared to the unadjusted. The non-zero coefficients in the overall survival model indicate that these variables are indeed strongly associated with survival outcomes (e.g. those with stage IIIb disease are likely to have better OS outcomes than those with stage IV disease); however, it is important to note that they do not indicate that these characteristics are imbalanced between treatment arms, which could suggest bias in the estimate of the treatment effect and therefore cannot be argued that we have overestimated the treatment effect for pemetrexed over placebo.

Indeed, there is no evidence to suggest the adjusted model offers more precision over the unadjusted model to the overall survival data of the treatment effect in PARAMOUNT (overall survival time ratio: adjusted model 1.21 95% confidence interval (CI): 1.02-1.44, unadjusted model 1.22 95% CI: 1.02-1.46) for the gamma distribution see Tables 3 and Table 4). This effect is indeed present not only for gamma but also for all the alternative parametric distributions tested in the original submission suggesting that both the unadjusted and adjusted models provide very similar if not equal estimated OS results and that this is consistent for all the parametric distributions.

We do not believe that appending an adjusted analysis to KM data is appropriate. Applying the adjusted risk equation requires assumptions regarding the proportion of patients with each characteristic (covariate) represented in the regression model. The ERG approach, using the adjusted analysis to extrapolate the unadjusted KM data, assumes that at the cut off point (20% of survival threshold), where the adjusted extrapolation begins, and the average patient characteristics remain the same as that at baseline. However, we know that, at the point of extrapolation, the proportion of patients with each characteristic will be different from baseline (e.g. the most severe patients (or the patients at highest risk of death) will already have died, resulting in a change to the average covariate values at this point of the KM data or indeed the level of significance).

We believe, therefore, that the use of the unadjusted model does not introduce any treatment effect bias and is therefore an appropriate approach to extrapolate our data beyond the trial duration. It is also worth noting that since the data at the end of the trial was very mature with only a small proportion of patients at risk of progression or death, the use of the full observed Kaplan Meier data and the unadjusted model better reflect the trial data.

Table 3 Overall Survival: estimates of the relative treatment effect pemetrexed plus BSC vs BSC alone unadjusted analysis

Distribution	HR/TR	p-value	95% LCI	95% UCI
Exponential	0.79	0.03	0.65	0.97
Weibull	0.77	0.01	0.62	0.94
Gompertz	0.78	0.02	0.63	0.95
Log-logistic	1.23	0.03	1.03	1.47
Lognormal	1.21	0.05	1.00	1.46
Gamma	1.22	0.03	1.02	1.46

Footnotes: TR – time ratio, HR – hazard ratio
LCI – lower bound confidence interval, UCI – upper bound confidence interval
NOTE: A unit change (PARAMOUNT+BSC vs. BSC) results in a decrease Hazard Ratio for the PH distributions indicating an increase in expected survival. A corresponding unit change in the non-PH distributions directly increases survival by this factor change (Ref STATA manual)

Table 4 Overall survival estimates of the relative treatment effect pemetrexed plus BSC vs BSC alone adjusted analysis

Distribution	HR/TR	p-value	95% LCI	95% UCI
Exponential	0.80	0.04	0.65	0.98
Weibull	0.77	0.01	0.62	0.95
Gompertz	0.78	0.02	0.63	0.96
Log-logistic	1.21	0.03	1.01	1.44
Lognormal	1.20	0.04	1.01	1.44
Gamma	1.21	0.03	1.02	1.44

Footnotes: TR – time ratio, HR – hazard ratio
LCI – lower bound confidence interval, UCI – upper bound confidence interval
NOTE: A unit change (PARAMOUNT+BSC vs. BSC) results in a decrease Hazard Ratio for the PH distributions indicating an increase in expected survival. A corresponding unit change in the non-PH distributions directly increases survival by this factor change. A corresponding unit change in the non-PH distributions directly increases survival by this factor change (Ref STATA manual)

2.b. Methodological assumptions: use of a case-mix adjusted utility model

The Committee and ERG's main concern with the use of utility values in the model refer to the fact that Lilly's basecase used an unadjusted utility value which was directly derived from the PARAMOUNT clinical trial rather than the preferred ERG option to use the adjusted utility regression model which would account for differences in patients' baseline characteristics.

The ERG has suggested that the adjusted utility model is probably more reliable than the unadjusted utility model. Although the two models provide different QALY values, the difference does not stem from the differences in patients' baseline characteristics. The reason for the different values comes as a result of the inclusion in the adjusted model of a "cycle" variable, that is to say a structural time variable. This variable was included to account for changes in utility values as the patient was moving through the cycles in the model. The introduction of this variable created some problems with the parameterisation of the adjusted model. Different parameterisation options were tested and only the quadratic function was statistically significant, however, this gave counterintuitive results from a clinical perspective as the adjusted model would provide higher utility estimates in the post-progression period than in the pre-progression one. Only when the cycle variable was incorporated as a linear function (which was statistically non significant) get obtained clinically plausible results. Due to these issues with parameterisation the variable "cycle" was not included in the basecase analysis.

Table 5 below shows that the unadjusted model provides a more conservative estimate than the adjusted model when the cycle variable is removed.

Table 5 Unadjusted versus adjusted Utility model (including and excluding "cycle" structural time variable)

	Unadjusted Model (Revised Basecase)	**Adjusted model *excluding* cycle variable**	**Adjusted model *including* cycle variable**
QALY	0.2480	0.2493	0.2392
ICER	£58,918	£58,608	£61,094

Adjusted model accounts for differences in: ECOG performance stage, historical illness, RECIST response, etc.)

References

1. Kleinbaum D. Survival Analysis: A Self-Learning Text (Statistics for Biology and Health). Springer; 2010.
2. Latimer N. NICE DSU Technical Support Document 14: Survival Analysis for Economic Evaluations Alongside Clinical Trials - Extrapolation with Patient-Level Data. SCHARR; 2011 Jun.
3. Patel K, Kay R, Rowell L. Comparing proportional hazards and accelerated failure time models: an application in influenza. Pharamceut Statist 2006;5:213-224.
4. Stein WD, Gulley JL, Schlom J, et al. Tumour regression and growth rates determined in five intramural NCI prostate cancer trials: the growth rate constant as an indicator of therapeutic efficacy. Clin Cancer Res. 2011 Feb 1517(4):907-17.
5. Stein WD, Huang H, Menefee M, et al. Other paradigms: growth rate constants and tumour burden determined using computed tomography data correlate strongly with the overall survival of patients with renal cell carcinoma Cancer J. 2009 Sep-Oct:15(5):441-7