Issues raised by GlaxoSmithKline and the Assessment Group response

Description of problem	Description of proposed amendment and result of amended model or expected impact on the result (if applicable)	Assessment Group response
1. Calculation error on the Calcs_Let sheet of the LRiG_executable_model_lapatinib+AI(061210) Cells AI48:AI77 on LET sheet, which represents the average PPS years by year post treatment initiation for LET patients, is incorrectly referencing the cells W163:W192 on the Calcs_ LapLet Sheet instead of referencing the same range on the Calcs_Let sheet.	Correct the calculation on the LET worksheet of the LRiG_executable_model_lapatinib+AI(061210). The amendment will generate a slightly increased ICER for the comparison of lapatinib plus letrozole versus letrozole. For a 20 year time horizon the estimated ICER will be £225,962 per QALY.	Agreed – this is a coding error
2. Error in Sampling the Decrement in Utility with Diarrhoea/Vomiting (D/V) for LAP+LET in the LRiG_executable_model_lapatinib+AI(061210) The model includes a calculation of the decrement in QALYs due to diarrhoea and vomiting (D/V) in LAP+LET patients. The default estimate of the decrement in QALYs appears to be calculated as the product of the estimated incidence of G3/4 D/V in the EGF30008 trial (8.21%) and the decrement in utility with D/V from the Lloyd study (0.0948). The resulting disutility is 0.0079 (see cell C82 of the parameters sheet). For the base-case, this calculation seems appropriate, although it assumes that patients are in the D/V state for an average of one year which may be overly conservative. Given the small disutility (<.01), it doesn't materially affect the conclusions. However, for the PSA, the model incorrectly samples the decrement in QALYs with D/V based on the disutility for D/V from the Lloyd study—that is, for the	Revise the calculation of sampling the decrement in utility for D/V with LAP+LET in the model The mean PSA value for QALYs with LAP+LET, the mean PSA value for incremental QALYs with LAP+LET, and the ratio of the mean incremental costs to the incremental QALYs will be similar to the base case estimates.	This problem was caused by the omission of one term in a control parameter formula in the 'Parameters' worksheet. In addition the Assessment Group has identified that the standard error for the D/V utility decrement was over-estimated. The result of correcting these problems is to yield a revised probabilistic ICER of £276,478 per QALY gained

Issues raised by GlaxoSmithKline and the Assessment Group response

Description of problem	Description of proposed amendment and result of amended model or expected impact on the result (if applicable)	Assessment Group response
PSA, the model fails to multiply the decrement in utility per person with the event (0.0948) by the proportion of patients with the event (0.0821). Accordingly, the mean decrement in QALYs from AEs in the PSA is approximately 0.0948 (This can be seen by taking the average of the values in cells Q7:Q1006 on the Rnums sheet). This (incorrect) decrement in QALYs due to D/V with LAP+LET largely offsets the gain in QALYs due to improved PFS with Lap+LET. Accordingly the mean incremental QALYs with LAP+LET in the PSA is underestimated by approximately 90% and the ratio of the mean incremental costs to the mean incremental QALYs is overestimated by a factor of 10. The resulting CEAC curves are accordingly severely biased/in error as is the scatter plot for the incremental QALYs vs. incremental costs. Also, the values reported in the second row of Table 28 of the report are also in error.		

Issues raised by GlaxoSmithKline and the Assessment Group response

Description of problem	Description of proposed amendment and result of amended model or expected impact on the result (if applicable)	Assessment Group response
3. Error in Inputs Used to Sampling the Proportion of Patients Progressing with LET in the LRiG_executable_model_lapatinib+AI(061210) The model adjusts PPS for LET using an adjustment factor of 0.953703704 calculated as 103/108 (see cells AK39:Al41 in Let sheet). For LAP+LET, the model uses a factor of 0.936936937 calculated as 104/111. For sampling the model uses beta distribution parameters of a=104 and b=7 for LAP+LET (a+b=111). However, for LET, the model uses beta distribution parameters of a=103 and b=3. To be consistent, the latter should use a=103 and b=5 (a+b=108). Note that this assumes that the values used to estimate the proportions are correct, although it is not clear how the 103 and 104 values were obtained from the EGF30008 trial data.	Check beta distribution parameters for sampling of the proportion of patients progressing. Using the value of b=5 for LET would likely reduced the mean PSA QALYs for LET and increase the mean PSA incremental QALYs for LAP+LET vs. LET and reduce the ratio of the mean PSA incremental costs to the mean PSA incremental QALYs. The magnitude of this effect is likely small however.	Agreed – this is a transcription error The result of correcting this problem in addition to changes made relating to Issue 2 is to yield a revised probabilistic ICER of £234,911 per QALY gained.

Description of problem	Description of proposed amendment and result of amended model or expected impact on the result (if applicable)	Assessment Group response
4. Potential Bias in the Sampling of PFS in the LRiG_executable_model_lapatinib+AI(061210) The model samples the PFS probability in each 28 day cycle by multiplying the base-case probability by normal random variable with mean of 1.0 and standard deviation equal to the standard error of the estimated mean PFS up to 504 days. Because PFS cannot exceed 1.0, the sampled values are truncated at 1.0. Because the PFS values for LAP+LET are closer to 1.0, they are more affected by this truncation. A more appropriate method for sampling would be to apply sampling to the sum of the PFS values up to 504 days rather than the survival probabilities. This would eliminate the need for truncating the distribution at 1.0	Revise the method for sampling PFS to avoid the potential bias associated with the truncation of sampled values at 1.0. The mean PSA value for QALYs for LAP+LET, the mean PSA incremental QALYs for LAP+LET vs. LET, and the ratio of mean PSA incremental costs to the mean incremental QALYs (labelled mean ICER) will be similar to the base case estimate.	The Assessment Group agrees that the discrepancy referred to will occur during PSA. It is not clear that the suggested alternative method would be effective, since sampling the total AUC up to 504 days would not allow reliable allocation of the variations from the expected mean within the overall period as is required to drive costs and outcomes in the model. In principle it might be possible to carry out pro-rata to the Kaplan-Meier point estimates at each event point in the period, but this is still problematic since standard software usually employs approximate symmetric confidence intervals, which are also inaccurate close to the baseline and subject to truncation. However, as similar uncertainty values are used in both arms, and the duration of exposure to differential truncation is small, it is considered that the net impact of the discrepancy arising from the current ratio method is minor and unlikely to impact the decision problem noticeably.
5. Lack of transparency which limits the testing of the robustness and reliability of the LRiG_executable_model_lapatinib+AI(061210) In general there is little information on the model inputs and results and this limits transparency and the testing of the model. Examples are as follows: a) The precise methods by which the two parameters of the exponential distributions for PFS after 504	For future consultations additional information on the Assessment Group model should be provided to facilitate testing.	a) PFS AUC up to 504 days was estimated by Kaplan-Meier analysis, censoring all remaining

Description of problem	Description of proposed amendment and result of amended model or expected impact on the result (if applicable)	Assessment Group response
days and for Post Progression Survival (PPS) were obtained using data from the EGF30008 trial is not in the model (or Assessment Report). The reason for using a 2-parameter exponential distribution (i.e., with the scaling factor) is also not described.		patients at 504 days. PFS modelling was carried out using the individual time to progression/censoring data for cases in both arms combined for time > 504 days, using non-linear regression to estimate both the exponential rate factor and the reference survival probability at 504 days. Similarly, PPS modelling used the individual time to death/censoring data for cases in both arms combined, using non-linear regression to estimate both the exponential rate factor and a significant non-zero intercept (presumably corresponding to an initial risk of very early mortality within a day or two of progression).
b) It would appear that in order to account for the proportion of patients who died prior to progression, the model adjusts expected PPS by a constant factor. This factor is 0.9537 for letrozole and 0.9369 for lapatinib plus letrozole. According to the model, these adjustments are calculated as the ratio of "progress patients" to "all patients" in each group and are103 / 108 for letrozole and 104 /111 for lapatinib plus letrozole) (refer to 'Let' cells AL39 to AL41 and 'LapLet' cells AO39 to AO41 of the lapatinib Assessment Group model). The source of this data is unclear.		b) The number of patients dying whilst in the pre-progression phase was calculated directly from the PPS, PFS and OS data provided to the Assessment Group in the clarification. Cross-matching of time of fatal events in the OS analysis with fatal events in the PPS analysis and progression events (fatal+non-fatal) in the PFS analysis allowed the number of fatal PFS events to be counted, as a basis for estimating the probability that patients will not be alive to accrue PPS time beyond progression.
c) For patients who progress during each 28 day cycle, the model calculates the number of PPS days that are accrued in each year following treatment initiation. It is not clear why this approach was employed, as the model assumes a constant risk of death given progression, and the need for what are		c) Modelling OS from the sum of PFS and PPS yields results in aggregate, but does not readily provide accurate allocation of PPS time between years, which is essential for discounting QALYs and also for assigning costs during PPS (BSC and terminal care). In general, no simple analytic

Description of problem	Description of proposed amendment and result of amended model or expected impact on the result (if applicable)	Assessment Group response
essentially “”tunnel states” for PPS is unclear.		formulae exist which allow this allocation to be carried out. Two spreadsheets (Calc_Let and Calc_LapLet) were developed to carry out this function, but using the number of patients progressing in each 28-day period as a cohort which was then subject to the risk of death in the PPS model, so that the time spent in each model year could be estimated. Because the method is subject to approximations, it was necessary to apply corrective adjustment factors to reconcile the totals to the PPS totals in the overall model.
d) It is difficult to assess the methods by which the PSA values were calculated. It appears that the model allows for a “standardized” or “unstandardized” PSA. The former uses values for selected parameters drawn from the parameter sets in the ‘RNums’ sheet of the model. The “Unstandardized” PSA uses sampling from normal distributions. It is not clear which approach was employed in the Assessment Report. It is not possible to determine the source of the values in the RNums table that are the basis of the “standardized” PSA.		d) One of several reasons for carrying out large numbers of model iterations in PSA is that the mean value of any uncertain parameter can deviate to an important degree from the long-run expected value, leading to unpredictable results. A simple mechanism for limiting this aspect of uncertainty and allowing some reduction in the number of iterations before stability is achieved is to pre-generate a set of random numbers for each independent variable such that the overall mean and standard deviation is less than a specified maximum limit. A similar approach can be used with correlated parameters constraining also the set of correlated values. This allows rapid production of PSA realistic results to aid development and exploration of interactions. The relative accuracy of these results can of course be verified by also running the model without constraints limited only by the time available.
e) There is a difference in the base case estimates given in the Assessment Report (£220,626 per QALY gained) and those shown in the model £215,504 per QALY gained (for the 20 year time		e) The amended values for utilities were introduced in response to a reviewer enquiry as to whether the higher rate of response to treatment in the EGF30008 (for lapatinib) compared to

Description of problem	Description of proposed amendment and result of amended model or expected impact on the result (if applicable)	Assessment Group response
horizon) for lapatinib plus letrozole versus letrozole. It appears that the reason for the difference between the results in the model and the Assessment Report relates to the utility values for PFS for lapatinib plus letrozole and letrozole. The Assessment Report states that these utilities are 0.7663 and 0.7623 for lapatinib plus letrozole and letrozole respectively. However, the model uses values of 0.779398257 and 0.774892291 respectively. The former are label as "original" in the model (see cells H79 and H80 on the Parameters sheet) whereas the latter are labeled as "revised" in the model (see cells I79 and I80 on the Parameters sheet). When the "original" values are used, the Model results match those in the Assessment Report. The source of the "revised" estimates is not provided, so it is impossible to ascertain which of the two sets of estimates is more appropriate. Also, it appears that the PSA is sampling based on the utility values used in the Assessment report not the revised values (see cells C41 and D41 on the Uncertainty sheet as well as the average values for cells G7:G1006 and H7:H10006 on the Rnums sheet). f) In general the model also incorrectly labels some cells as "utilities" that should be more appropriately labeled QALYs as they represent the product of utility values and life years. Improving the labeling/documentation of the model might facilitate the identification of calculation errors such as that described above.		TAnDEM (trastuzumab) had been fully reflected in the two models (i.e. were the utility values comparable). The original values had involved an adjustment for the duration of response derived from the TAnDEM trial, and re-used in the lapatinib model for lack of equivalent data from EGF30008. However, a modification was introduced which explicitly recognised the higher response rates in both arms of EGF30008, and this led to increased PFS utility estimates for use in the lapatinib model. f) Agreed

Issues raised by GlaxoSmithKline and the Assessment Group response

ASSESSMENT GROUP SUMMARY

The combined effect of applying corrections/amendments to the Assessment Group model in respect of Issues 1, 2 and 3, together with Issue 5(e) are to reduce the deterministic ICER for lapatinib+letrozole vs. letrozole to £225,962 per QALY gained. The revised probabilistic ICER is £228,913 per QALY gained with no measurable probability that lapatinib is cost-effective at a willingness-to-pay threshold of less than £50,000 per QALY

Issues raised by Roche and the Assessment Group response

Description of problem	Description of proposed amendment and result of amended model or expected impact on the result (if applicable)	Assessment Group response
1. Adjustment for patients who died in PFS on the calculation of post-progression survival In the ACD section 4.2.15, it is explained that in the Assessment Group model, "the estimate for overall survival was obtained by combining estimates of mean progression-free survival and mean post-progression survival in each group, and adjusting for the patients who died at or before progression (5.8% in the anastrozole alone group and 9.3% in the trastuzumab plus anastrozole group)". In the TAnDEM trial, 6 patients in each treatment arm died at or before disease progression. Given that there were 103 patients in the intervention arm and 104 patients in the control arm, it would appear that the anastrozole figure was calculated correctly (5.8%) but the proportion in the trastuzumab arm has been overestimated (97/103 = 5.8%).	On Sheet TA_AI, cell AO41, which we believe is incorrectly labeled "Progressed AI only patients" (as this refers to the trastuzumab arm, it should probably read "Progressed TR+AI patients", the current value is 93. This value should be 97 if the intent is to reflect the number of patient who did not die before or during disease progression (i.e. 103 patients started the trial and 6 of these patients died). When changing the incorrect cell within the Assessment Group model (Sheet TA_AI, cell AO41), this results in a decrease to the ICER (**from £73,135 to £69,514**) resulting from an improved mean time in PPS determined from the trastuzumab arm.	The number of TAnDEM patients in each arm of the trial who died prior to or at disease progression was not stated explicitly in any of the documentation provided by the manufacturer (NICE submission, CSR or response to clarification requests). In addition none of the documents provided a fully analysed table or chart showing the disposition of patients by nature and timing/phase in the patient pathway, from which the number of deaths could have been deduced. However, in retrospect the Assessment Group acknowledges that it is possible to derive the figures referred to by the manufacturer from the data supplied in response to clarification questions. The Assessment Group confirms that the impact of this correction on the economic results is as stated by the manufacturer (Issue 2, below: 20 year ICER reduced **from £73,135 to £69,514**/QALY gained).

Issues raised by Roche and the Assessment Group response

Description of problem	Description of proposed amendment and result of amended model or expected impact on the result (if applicable)	Assessment Group response
2. Calculation of Progression-Free Survival and the associated trastuzumab costs The methods used in deriving the PFS Weibull curve were not clearly presented and it is unclear as to whether estimates of uncertainty around their parameter estimates were incorporated into the PSA performed. From comparison of the Roche and Assessment Group models, it is clear that the estimate of time in PFS differs considerably between the Roche original model (1.19 years) and the assessment group's model (1.30 years). In the Roche model, we have utilized the Kaplan Meier PFS curves for both the trastuzumab arm and anastrozole arm from the TAnDEM trial. This data was complete (i.e. no patients remained in PFS at the end of follow-up) and therefore no extrapolation was conducted. It can be considered that the mean time in PFS presented in our model reflects the mean time observed in the clinical trial.	Utilise the KM curves to determine the time in PFS and the Roche methodology for utilising actual individual patient weights from the clinical trial to determine the required number of vials of trastuzumab. The impact of this potential overestimation of time in PFS in the Assessment Group model will affect multiple parameters in the model, but the most considerable impact of this overestimate as it relates to the ICER will likely be the overestimation of the cost of trastuzumab and anastrozole combination therapy which is given until disease progression. The cost of the trastuzumab and anastrozole combination therapy is estimated to be approximately £6,500 greater in the Assessment Group model (£31,272) compared to the Roche model (£24,774). In order to provide a crude calculation of the impact of potentially overestimating average time in PFS and therefore overestimating the cost of trastuzumab by £6,500, we [Roche] have removed an additional £6,500 from the numerator of the Assessment Group modified ICER, resulting in a downwards shift of the ICER **from £69,514 to £57,591**.	The manufacturer comments with some justification on the use of projective modelling, rather than Kaplan-Meier analysis of the full data set to estimate the difference in PFS. Although in general the Assessment Group supports the position that trial data should normally be preferred to model estimates, the TAnDEM trial data raised some particular questions of uncertainty. The most obvious concerned the treatment of the last (censored) patient in the AI arm who remained alive and progression-free until 1900 days – nearly twice as long as the last progressed patient (980 days). By contrast the trastuzumab arm data included almost equal numbers of censored and progressed patients over 1000 days ending with a progression event. Several alternative rules for handling these data were considered, but these pointed to quite different estimates of incremental PFS gain indicating wider uncertainty than that represented by conventional Kaplan-Meier analysis. As a result the Assessment Group made a judgement that employing projective modelling to both arms in this particular case was more likely to reflect the uncertainty in outcome estimation than any of the alternatives. This is not ideal, but was dictated by the trial data with extended and uneven distribution 'tails' in a trial with small patient numbers. The Assessment Group is not aware of any general theoretical work which might give clear grounds for choosing one approach over the other.

Issues raised by Roche and the Assessment Group response

ASSESSMENT GROUP SUMMARY

A correction to the AG model reducing the ICER to £69,514 per QALY gained. The AG does not accept the superiority of K-M estimation over projective modelling (which would reduce the ICER further to about £58,000/QALY) but recognises that this is a debatable issue.