Response to the Evaluation Report for Lapatinib for HER2 over-expressing breast cancer (second ACD consultation)

GlaxoSmithKline 4 November 2008

GSK welcomes the opportunity to respond to the Evaluation Report, in particular the Decision Support Unit Report and Addendum dated 7 September 2008.

Our response is structured to reflect the individual sections in the DSU report

1. Section 2.1. DSU Report: Trial EGF100151

Issue

Trial EGF100151 (lapatinib and capecitabine compared to capecitabine alone) – Updated data (September 07 cut-off) from EGF100151 were only presented for overall survival

Response

The updated overall survival analysis was conducted at the request of the CHMP as part of the regulatory process for lapatinib. An updated analysis of PFS /TTP was not conducted because of the major practical challenges associated with collecting follow-up data at this stage of a clinical trial (i.e. after it had closed). Such an update would have had to have been limited to an investigator assessment of tumour progression (which was not a primary endpoint) because an updated assessment by independent reviewers had not originally been planned and would take approximately 12-24 weeks to conduct. This was not feasible within the timeframe GSK was given to respond to the CHMP's questions.

In addition, an update would have introduced potential for bias:

- 33 patients on the capecitabine only arm crossed over to receive lapatinib in combination with capecitabine following the halting of the study on 03 April 2006. Twenty nine of these patients crossed over prior to their disease progressing.
- Following the Independent Data Monitoring Committee recommendations, study investigators had been made aware of the interim study results. As it was an open-label study, knowledge of these interim results may have had the potential to introduce bias in the investigators assessment of progression/ interpretation of the scans. This could also have led to investigators identifying a progression event which then could not be confirmed by independent review.

Overall, GSK believed such an update to be flawed, and that it would not provide any useful data beyond that of the April 2006 analysis.

2. Section 2.4. DSU Report: Pooled estimate of trastuzumab efficacy

Issue

The DSU states that "*It should be noted that a standard Fixed or Random effects meta-analysis approach produces implied hazard ratios which are considerably larger (that is, less favourable for trastuzumab) than the weighted method adopted*."

Response

Our primary analysis was based on estimates of effectiveness of trastuzumab obtained from the GBG 26 / BIG 3-5 study. The analysis based on pooled results of this and uncontrolled studies was included as supporting evidence with the understanding of its limitations; in particular the breaking of randomisation.

With regard to the methods used to pool results from the various trials, we recognize that fixed or random-effects meta analysis may be more appropriate methods for pooling of estimates across multiple studies than weighting results of individual studies by the number of subjects. However, few of the studies in the analysis reported estimates of the variance of the median TTP or OS (i.e., standard deviations or errors, confidence intervals). Accordingly, calculation of fixed or random effects estimates of the median TTP or OS was not deemed feasible.

Nevertheless, assuming that the variance of the median TTP or OS is proportionate to the median, studies with lower median TTP or OS would have lower variances (controlling for sample size), and therefore would be given greater weight in fixed (and potentially random-effects) meta analysis (because studies are weighted by the inverse of the variance of the median), resulting in a larger (less favourable) implied HR for trastuzumab. Similarly, because random-effects meta analysis gives relatively equal weight to studies regardless of sample size (because studies are weighted by the inverse of the sum of the within and between trial variance), and because the larger studies tended to have higher median TTP or OR, random-effects meta analysis also would give relatively greater weight to studies with smaller TTP or OS, again resulting in a larger (less favourable) implied HR for trastuzumab. The use of the weighted mean median for TTP and OS was both appropriate and conservative.

3. <u>Section 2.4. DSU Report: Pooled estimate of trastuzumab efficacy</u>

Issue

The DSU point out that an increase in the hazard ratio for trastuzumab regimens in relation to capecitabine (i.e. a decrease in their relative effectiveness versus the lapatinib combination) will favour the cost effectiveness of lapatinib.

Response

This assumption is incorrect, because the decreased time period over which patients would be receiving trastuzumab, with associated lower costs, has a paradoxical impact in decreasing lapatinib cost effectiveness (i.e. increasing the incremental cost effectiveness ratio, ICER), and vice versa, when compared with trastuzumab regimens. This relationship between the costs and effects of lapatinib and trastuzumab regimens is likely to apply over the range of plausible values for these parameters.

4. Section 2.4. DSU Report: Pooled estimate of trastuzumab efficacy

Issue

In Section 2.4 of the DSU Report, it is stated

"An alternative approach, which would have taken into account the potential for bias in comparisons derived from single arm studies, is that of Begg & Pilote (1991)[9], ….."

Response

We recognize that the approach proposed by Begg & Pilote represents a possible alternative that would use the information from the GBG 26 / BIG 3-05 study as well as the uncontrolled studies. However, as with fixed- and random-effects meta analyses, the method of Begg & Pilote requires estimates of the variance of the median TTP or OS for each study, which was largely unavailable. Use of this method therefore was not feasible.

5. Section 2.4. DSU Report: Estimation of hazard ratio for trastuzumab plus capecitabine compared with capecitabine monotherapy from GBG26/BIG3-05 Study

Issue

In Section 2.4 of the DSU Report, states:

"An alternative approach, which would have taken into account the potential for bias in comparisons derived from single arm studies, is that of Begg & Pilote (1991)[9], though a more elegant approach would have been to conduct a Mixed Treatment Comparison (MTC), in which single arm studies are included via a sensitivity analysis with potential adjustment for bias, though it is accepted that this represents a step forward in the manner in which MTC methods have otherwise been applied to date.

Using MTC methods to obtain a hazard ratio HR for trastuzumab plus capecitabine compared to lapatinib plus capecitabine using the results from the Cox model reported in the trial, the hazard ratio for TTP/PFS is 0.96 (95% CI: 0.64 to 1.45) and for OS is 0.84 (95% CI: 0.50 to 1.43). However if the data from the Weibull approximated results are used in the MTC analysis, the results are: HR for TTP/PFS is 1.22 (95% CI: 0.89 to 1.67); and HR for OS is 1.04 (95% CI: 0.76 to 1.44)."

Response

Whilst we agree with the DSU that the MTC represents a potential alternative approach for estimating the effects of T+C and L+C on TTP/PFS and OS compared with C-only, and may be especially useful when there are multiple (i.e., greater than two) comparators and studies to evaluate, given that there was only one other comparative trial to be included in the analysis (i.e., the GBG 25 / BIG 3-05 study), it was felt that the use of MTC in this instance was unnecessary. A comparison of the HRs for TTP/PFS and OS with T+C vs. L+C obtained by the DSU using MTC versus those implied by dividing the HRs for T+C vs. C-only by that for L+C vs. C-only as obtained directly suggests that the use of MTC has little impact on estimates of effectiveness.

Table 1. Comparison of estimated HR for T+C vs L+C from DSU MTC vs. ratio of HRs for T+C vs. C-only from GBG 26 / BIG 3-05 vs. L+C vs. C-only from EGF100151

Outcome	HR from AFT Weibull Models		$HR^{T+C \text{ vs } L+C}$	
	$HR_{T+C \text{ vs C-Only}}$ (GBG 26 / BIG 3-05)	$HR_{L+C \text{ vs C-only}}$ (EGF100151)	$HR_{T+C \text{ vs C-only}} / HR_{L+C \text{ vs C-only}}$	MTC
TTP/PFS	0.7397	0.6085	1.22	1.22
OS	0.8696	0.8703	1.00	1.04

*HRs from AFT Weibull models

6. Section 2.4. DSU Report: Estimation of hazard ratio for trastuzumab plus capecitabine compared with capecitabine monotherapy from GBG26/BIG3-05 Study

Issue

Regarding the use of AFT regression to estimate the HR for T+C vs. C-only instead of using the hazard ratio reported in the abstract and poster, the DSU report states:

"It is also unclear why an Accelerated Failure Time (AFT) Weibull model was used to estimate the hazard ratios when these were available directly from the abstract. It should also be noted that the hazard ratios using the Weibull model show a larger effect than those reported in the abstract"

Response

We sought to derive estimated HRs for TTP/PFS and OS for T+C vs. C-only using data from the GBG 26 / BIG 3-05 study that would be comparable to those estimated for L+C vs. C-only using data from the EGF100151 study. We estimated the HR for L+C vs. C-only by fitting Weibull survival functions to patient-level failure-time data on PFS and OS from the EGF100151 trial using Accelerated Failure Time (AFT) regression (SAS PROC LIFEREG). This approach uses maximum likelihood estimation to estimate the three parameters of the Weibull models (λ, γ, and $HR_{\text{Treatment vs. Control}}$). Whilst this approach yields a HR for treatment versus control, the HR obtained from the Weibull AFT regression model is not directly comparable to the HR obtained from Cox proportional hazard regression models (which is likely to be what was reported in the GBG 26/BIG 3-05 abstracts and posters, as per accepted conventions for reporting hazard ratios in clinical trials).

For example, as shown in Table 2 below, the HR for PFS for L+C vs. C-only in EGF100151 obtained from the Cox regression was 0.55 whereas that obtained from the AFT Weibull regression was 0.6085. Similarly, the HR for OS for L+C vs. C-only in EGF100151 from the Cox regression was 0.82 whereas that from the AFT Weibull regression was 0.8703. As the HRs from the Cox regressions are consistently lower (i.e., more favourable) than those based on AFT Weibull regressions, use of the HRs for L+C vs C-only obtained from the Weibull model along with the HRs for T+C vs. C-only as reported in the GBG 26 / BIG 3-05 SABCS 2007 and/or ASCO 2008 posters might bias clinical effectiveness results in favour of T+C. and therefore cost effectiveness would be biased in favour of L+C, for reasons pointed out in point no. 3 above.

Table 2. Comparison of HRs for investigator assessed PFS and OS for L+C vs. C-only in EGF100151 (Sep2007 data cut-off) using Cox proportional hazards regression and AFT Weibull regression

	Cox Regression		AFT Weibull Regression	
Outcome	**HR vs C-only**	**95%CI**	**HR vs C-only**	**95%CI**
L+C (EFG100151)				
PFS	0.55	0.41-0.74	0.6085	0.4635 - 0.7496
OS*	0.82	0.65-1.04	0.8703	0.7357 - 1.0295
GBG 26 / BIG 3-05†				
TTP	0.71 / 0.69	n/a	0.7397	0.5559 - 0.9842
OS	079 / 0.76	n/a	0.8696	0.6651 – 1.1370

*In these analyses, patients who crossed over from C-only to L+C were censored at date of cross-over.
†SABCS 2007 / ASCO 2008

To ensure comparability of survival function estimates for L+C and T+C, and to avoid biasing in favour of L+C cost effectiveness, we estimated HRs for T+C vs. C-only in GBG 26 / BIG 3-05 using the same methods as we used to estimate HRs for L+C vs. C-only in EGF100151. Specifically, we estimated the three parameters of PH Weibull models for T+C and C-only for TTP (λ^{TTP}, γ^{TTP}, $HR^{TTP}_{T+C\ vs\ C\text{-}only}$) and OS ($\lambda^{OS}$, γ^{OS}, $HR^{OS}_{T+C\ vs\ C\text{-}only}$) in GBG 26 / BIG 3-05 using AFT regression and product-limit survival estimates for TTP and OS reported at ASCO 2008.

As shown in Table 2 above, the HRs for T+C vs. C-only in GBG 26 / BIG 3-05 obtained by this method are higher than those reported in the SABCS and ASCO poster based on Cox regression (contrary to the DSU statement that "*the hazard ratios using the Weibull model show a larger effect than those reported in the abstract")*. This result is consistent with the result obtained for the HR for L+C vs. C-only in the EGF100151 trial, and suggest use of the AFT regression model was appropriate and necessary to obtain unbiased estimates of the relative clinical and cost effectiveness of T+C vs. C-only.

7. Section 2.4. DSU Report: Estimation of hazard ratio for trastuzumab plus capecitabine compared to capecitabine monotherapy from GBG26/BIG3-05 Study

Issue

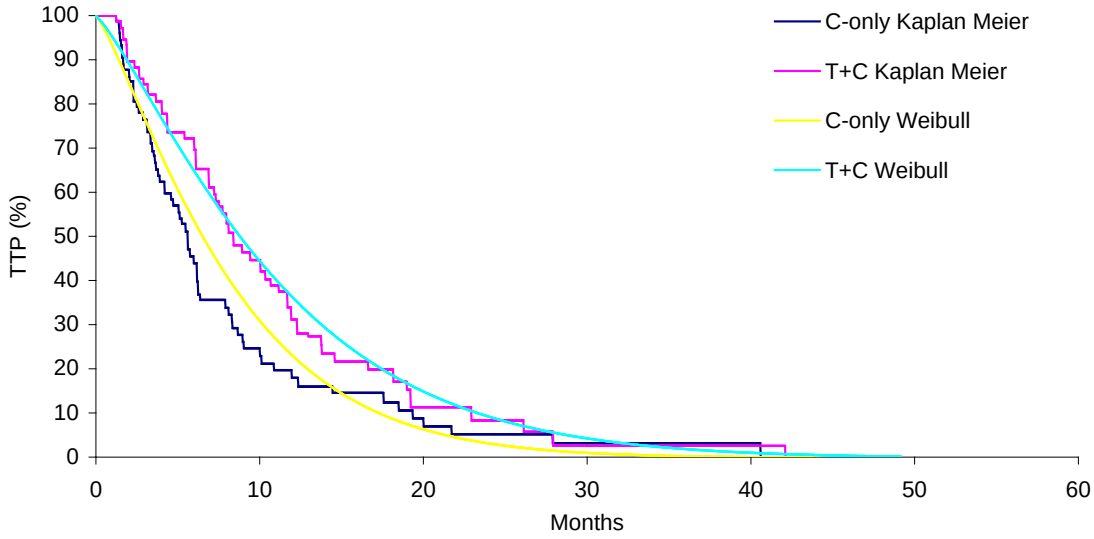
Section 2.4 of the DSU report states:

"The efficacy estimates for trastuzumab-containing regimens that were used in the economic analysis were based on a reanalysis of the data from the GBG 26/BIG 3-05 study, based on a conference abstract/poster, whereby the Kaplan-Meier curves were digitized and a Weibull distribution fitted for both TTP and OS. Both however show considerable lack-of-fit for the capecitabine-only group (GSK appendix 4, figures 5 and 6)."

Response

The DSU is correct that that the AFT PH Weibull estimated TTP for C-only was greater than the empirical survival distribution estimates between approximately 3 and 15 months of follow-up, and that this may have biased the Weibull estimates in favor of C-only and against T+C (Figure 1 overpage).

Figure 1. Kaplan-Meier and Weibull estimated TTP from GBG 26 / BIG 3-05



No. at Risk						
T+C	74	15	5	2	1	0
C-only	77	29	4	1	1	0

This divergence may reflect the failure of the model to account for a minimum failure time, reflected in the absence of TTP events during the first few weeks of the trial as a consequence of the absence of scheduled assessments. However, the Weibull estimate TTP was lower than the empirical estimate after 15 months, offsetting the higher estimates during the prior period. This is reflected in the fact that the difference in the "area under the curve" (AUC) for TTP for T+C vs C-only was one week larger based on the Weibull model (12.3) than the Kaplan Meier estimates (11.7 weeks) (Table 3).

Table 3. Comparison of Kaplan-Meier and Weibull model estimated "area under the curve" for TTP and OS for T+C and C-only in GBG 26 / BIG 3-05

	Kaplan-Meier			Weibull		
	T+C	**C-only**	**Difference**	**T+C**	**C-only**	**Difference**
PFS	47.3	35.6	11.7	48.1	35.8	12.3
OS	105.7	96.8	8.9	107.0	95.9	11.1

Area under curve calculated as sum of survival function estimates to 42.3 months

To assess the potential effects on our estimates of the timing of TTP assessments, we refit the Weibull model for TTP in GBG 26 / BIG 3-05, subtracting from each patient's failure time the minimum failure or censor time observed in the trial (32 days), thus shifting the survival distribution to the left and eliminating the period when no events occurred.

Using this approach, the HR for T+C vs C-only was estimated to be 0.6843. While this estimate is lower than that obtained for T+C vs. C-only using the raw data (0.7397), it is still greater than that for L+C vs C-only (0.60847). Replacing our revised base-case estimate of the HR for T+C vs. C-only with this estimate, and setting all other parameters to those employed in Scenario 9 (see GS@'s response to the first ACD), the model yields the results in Table 4 below.

Table 4. Base-case results-Scenario 9 plus revised HR for T+C.

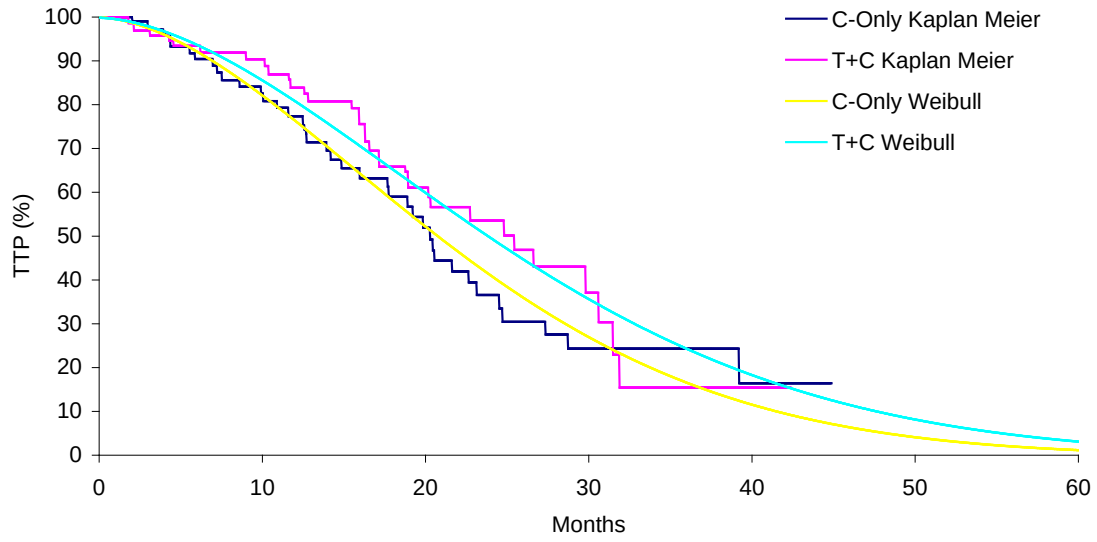
Outcome	L+C	Difference vs. L+C					
		C-Only	T+C	T+V	Usual Care	V-only	T-only
Total Costs, disc.	26,939	14,015	-2,251	-4,992	4,166	11,726	-424
QALYs	0.932	0.158	0.016	0.016	0.079	0.158	0.016
Cost per QALY		93,825	Dominant	Dominant	55,833	78,503	Dominant

New lapatinib price (£11.49), Sep2007 OS from EGF100151, HRs for T+C for PFS and OS from GBG 26/BIG 3-05 with failure times adjusted to account for minimum failure time, 88.4% TZ 6 mg/kg/d q3w and 11.6% TZ 2 mg/kg/d q1w, 15% wastage with TZ.

As these results show, the use of the alternative estimate has little qualitative effect on results.

Regarding OS, the Weibull curve for C-only is greater than the Kaplan Meier curve for C-only between 20 and 30 months (Figure 2). For T+C, the Kaplan-Meier curve is generally greater than the Weibull curve between approximately 7 and 30 months, but substantially less than the Weibull curve thereafter.

Figure 2. Kaplan-Meier and Weibull estimated OS from GBG 26 / BIG 3-05



No. at Risk						
T+C	74	50	21	8	2	0
C-only	77	59	27	6	1	0

Unlike the difference in TTP, this difference does not appear to be a consequence of the timing of scheduled assessments, rather it is due to the flattening of the Kaplan-Meier OS curve for C-only at approximately 28 months and the sharp drop in OS for T+C at approximately 30 months, resulting in the two curves crossing at approximately 31 months. As with TTP, the difference in the AUC for T+C vs. C-only using the Weibull model is greater than that based on the Kaplan-Meier estimates (11.1 vs. 8.9 wks) (Table 3).

8. Section 3.1.8. DSU Report: Comments on amendments to the economic model

Issue

The DSU report states:

"In particular, one of the key assumptions in the original model was that the length of survival post-disease progression would be the same for lapatinib/capecitabine and the trastuzumab containing regimens. This has been amended in the updated analysis so that the PPS of patients treated with trastuzumab containing regimens is longer than that for those treated with lapatinib/capecitabine."

Response

The assumption of equal PPS for L+C and trastuzumab-containing regimens in the original analysis was made because of the lack of robust data on OS for trastuzumab containing regimens. Given the availability of data on OS for T+C and C-only from the GBG 26 / BIG 3-05 trial, the assumption of equal PPS was no longer required. In the updated analysis, we used to data from the GBG 26 / BIG 3-05 trial to estimate the HR for OS with T+C vs C-only directly. PPS was then calculated as the difference between OS and PFS (as was done for L+C and C-only). In the revised analysis, the PPS with T+C is greater than that with L+C, which is conservative with respect to the benefits of lapatinib treatment.

9. Section 3.2.3 Market research data on the use of trastuzumab post progression in the metastatic setting

Issue

The DSU commented on a lack of clarity from the Cegedim Dendrite market research results as to whether clinicians had the opportunity to retrospectively review their patients records, or if information was based solely on clinician recall.

Response

The potential limitations of this research are explicitly included in Appendix 1 section A 1.1.2 in GSK's response to the first ACD. Here the level of evidence is described as being lower than the IMS data as the study is based on clinicians' perceptions rather than a review of patient records. However similarities in the results of these two data sets should provide an increased level of confidence around these data.

10. Section 3.2.3 Market research data on the use of trastuzumab post progression in the metastatic setting

Issue

The DSU commented that full details of data collection, including the methods of recruitment of respondents, characteristics of respondents / non-respondents and response rates were not available for any of the market research data therefore it was not possible to single out a specific source of data as being superior to the others.

Response

We strongly believe that the IMS Oncology Analyzer is the most reliable data source for the following reasons: IMS Oncology Analyzer uses a representative panel of hospitals which are geographically varied, and which include a minimum of 70% of all major cancer centres. IMS data is a longitudinal database enabling full patient history to be obtained since diagnosis. Robust measures of accuracy and quality control are also routine within this type of database. The more recent data presented in table 1 of the second ACD response is also representative of around 20% of the total metastatic breast cancer population.

11. <u>Section 3.2.3 Market research data on the use of trastuzumab post progression in the metastatic setting</u>

Issue

The DSU point out that *"there is variation between the sources in the estimated proportion of patients receiving trastuzumab containing regimens after progression in the metastatic setting, most notably between the data submitted by GSK and Roche*".

Response

GSK is concerned about the amount of emphasis being given to the data provided by Roche. No methodological detail for the market research has been provided and as acknowledged by the Committee in section 4.3 of the ACD the clinical specialist advisers considered the higher estimates of between 49% - 56% to be more appropriate. In addition to the IMS data, the Cegedim Dendrite data gives similar results and is a reliable and widely recognized data source for the following reasons: A mailing list was generated using the published Cancer Care 2007 National Cancer Directory, from which all oncologists treating in both NHS and private centres were selected, thereby ensuring a random selection of participants. Almost 600 senior oncologists across the UK were mailed and invited to participate, over a period of eight weeks. Explicit consent was obtained from all responders to have their names released with the research purpose fully outlined in the invitation process. The data were entered externally by an independent provider with quality control, cross checking and participant identification matching being carried out.

12. <u>Section 3.2.3 Market research data on the use of trastuzumab post progression in the metastatic setting</u>

Issue

The first ACD and ERG report noted that "*there was a slight bias towards the Greater London region, but that this was unsurprising due to the relatively larger population of patients and clinicians in this geography. Therefore the MBC-OA-ETS* [IMS Oncology Analyzer] *data should not be seen as being over representative of any particular region, or any type of hospital".*

Response

It may be of interest to note that the average mortality rate for women with breast cancer (a reasonable proxy for metastatic breast cancer prevalence) is 336/100,000 of population, versus 312/100,000 for non-London regions. It is not unreasonable to assume that the slight weighting towards London in the IMS database is reflective of the higher apparent prevalence of breast cancer in this area.