National Institute for Health and Care Excellence

Centre for Health Technology Evaluation

Pro-forma Response

ERG report

Aflibercept for the treatment of metastatic colorectal cancer which has progressed following prior oxaliplatin-based chemotherapy

You are asked to check the ERG report from the NHS Centre for Reviews and Dissemination, University of York to ensure there are no factual inaccuracies contained within it.

If you do identify any factual inaccuracies you must inform NICE by **5pm** on **6 May 2013** using the below proforma comments table. All factual errors will be highlighted in a report and presented to the Appraisal Committee and will subsequently be published on the NICE website with the Evaluation report.

The proforma document should act as a method of detailing any inaccuracies found and how and why they should be corrected.

ISSUE 1 Difference between treatment groups in subgroup of patients with liver metastases

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
[Factual error] **Page 14, paragraph 1** **Page 59, second paragraph of section 4.3** **Page 74 paragraph 1:** "Subgroup analyses suggest that patients with less advanced disease (ECOG PS 0, number of organs with metastases 1 or less, and liver only metastases) may be more likely to benefit from aflibercept, although there was no statistically significant interaction."	We suggest adding some text to clarify that there was a significant treatment effect observed in the liver metastases only group, both in terms of OS and PFS. The following details need to be incorporated: The pre-specified subgroup analyses of patients with liver metastases showed a greater OS benefit with aflibercept than placebo, and a test for interaction was significant at the 10% level (P = 0.0899). Furthermore, the pre-specified subgroup analyses of patients with liver metastases showed a greater PFS benefit with aflibercept than placebo, and a test for interaction was significant at the 10% level ([redacted]).	This is a factual error due to over-generalisation. It may mislead the AC in respect of the efficacy of aflibercept plus FOLFIRI in the liver metastases only subgroup.	This is not a factual inaccuracy. The test for interaction was significant at the 10% level, rather than the generally accepted 5% level. However, this has been amended for clarity.

ISSUE 2 Comparison between VELOUR trial and UK population

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
[Factual error / inconsistency] **Page 14, first paragraph** " and the proportion of patients who had metastatic involvement of only one organ was higher in the trial population than in patients seen in standard UK practice." **Page 35, fourth paragraph** "…and that the proportion of patients who had metastatic involvement of only one organ (42-44% of the trial population) was also higher in the trial population than in patients seen in standard UK practice."	We suggest removing statements that the proportion of patients who had metastatic involvement of only one organ was higher in the VELOUR trial population than in patients seen in standard UK practice.	The statements made in relation to patients in current UK practice and levels of organ involvement are not referenced and no data are provided for comparison to justify the statements made. In our submission we explicitly compared data from observational research of UK clinical practice and found that in a real-life clinical setting, [redacted] of the population had metastatic involvement confined to one organ (see Table B16 of MS), This is higher than the proportion in VELOUR (41.8-44.1%) and quite contrary to the statement in the report.	Clinical advice to the ERG indicated that these statements were true. However, the ERG accepts that as they could not be substantiated by referenced material, it should be made clearer that this statement is based on clinical experience only. This has been incorporated into the erratum.

Page 50, last sentence of last paragraph: "…patients seen in practice are more likely to have more than one site of metastatic disease, to be older and have a worse ECOG PS than patients in the VELOUR trial."			

ISSUE 3 Comparison between VELOUR trial and UK population – age and fitness

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
[Factual error / inconsistency] **Page 13 final paragraph** "However, the demographic and disease characteristics of the trial participants suggests that they were potentially younger and healthier than the population eligible for aflibercept plus FOLFIRI in UK clinical practice." **Page 35 paragraph 4** "therefore, the patients included in the trial were younger than those seen in UK clinical practice". **Page 50, paragraph 4** "However, differences in characteristics between trial participants and patients seen in standard UK practice suggest that patients in the VELOUR trial are likely to be fitter than patients eligible for aflibercept plus FOLFIRI in practice; patients seen in practice are more likely to have more than one site of metastatic disease, to be older and have a worse ECOG PS than patients in the	We suggest removing altogether these sentences stating that the VELOUR trial participants are potentially younger and healthier than the population eligible for aflibercept plus FOLFIRI in UK clinical practice.	The statements questioning the generalisability of the VELOUR patients to those likely to receive aflibercept in UK practice appears to have been made on the basis of the broadest mCRC population, and not the specific second-line chemotherapy-treated population. The population eligible for aflibercept plus FOLFIRI in UK clinical practice represents a subset of all mCRC patients; namely those patients who are fit enough to tolerate second-line combination chemotherapy. As such, these patients are highly likely to be younger and fitter than the mCRC population as a whole. Indeed, in the observational RU study, which included all patients treated with an irinotecan-based regimen following an oxaliplatin-based regimen (i.e. a patient group who could be considered potentially eligible for aflibercept) and was conducted in 4 major UK hospitals, the median age was 63, which is similar to that in VELOUR. [redacted]	Clinical advice to the ERG indicated that these statements were true. However, the ERG accepts that as they could not be substantiated by referenced material, it should be made clearer that this statement is based on clinical experience only. This has been incorporated into the erratum.

VELOUR trial." **Page 59 paragraph 4** "The demographic and disease characteristics of the trial participants suggest that they were potentially younger and healthier than the population eligible for aflibercept plus FOLFIRI in UK". **Page 74 paragraph 1** "..characteristics of the trial participants suggest that they were potentially younger, with comparatively better performance status than the population likely to be eligible to receive aflibercept + FOLFIRI within UK routine clinical practice. In practice."		[redacted].	

ISSUE 4 Details of the Patient Access Scheme

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
[Inconsistency inviting confusion] **Page 15 paragraph 1** "Despite repeated requests from the ERG for further information on the nature of the PAS, insufficient information was provided to the ERG to assess whether the implementation of the PAS in the economic model was done correctly". This also relates to: **Page 17, paragraph 4** **Page 89 paragraph 1** **Page 131, paragraph 2**	We suggest adding a statement in the relevant sections of the document to indicate that the PAS is a simple discount-based scheme which has been appropriately applied. The statements regarding repeated requests for information on the PAS should be removed.	The PAS is a simple discount-based PAS; the price in the model is the agreed PAS price, and was confirmed before the issuance of the ERG report. Therefore there should be no question as to whether the "implementation" was correct or not. Indeed, the ERG acknowledge on page 89 that - assuming the PAS price has been agreed - then the PAS has been applied correctly. With this repeated expression of concern throughout the document, the reader could be forgiven for thinking this is an area of significant uncertainty – it is not. Finally, no requests for this information were received by the manufacturer.	This is not a factual inaccuracy. The ERG was only informed that the PAS was a simple discount over the list price of aflibercept, and that the PAS had been approved by the Department of Health. Nevertheless, no further information regarding the actual size of the discount was provided, and therefore the ERG could not verify whether the prices applied in the model resulted from the correct implementation of the PAS.

ISSUE 5 Assumption of increasing treatment effect over time

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
[Factual errors] **Page 15 of ERG report, end of last paragraph** "Implicitly the use of independent parametric functions resulted in an increasing hazard ratio of OS over time. That is, the relative effectiveness measure applied vs. FOLFIRI alone was more favourable as the time horizon increased"	Removal of the sentence on page 15 Amendment of the text on page 76-77 to reflect the fact that in the manufacturer's base case analysis, the treatment effect initially improves during the trial-follow-up period and then diminishes for later time periods. Amendment of the text on page 109 to "Fitting separate parametric log-logistic curves to each treatment arm	The reports' statements are incorrect. The OS functions used for the base-case analysis in the MS predicted an increasing treatment effect during the first year, but thereafter a decreased treatment effect is modelled over time.	**ERG correction:** **Page 15, last paragraph** **<u>The use of the independent parametric functions resulted in the hazard ratio decreasing from HR>1 to HR<1 over the first year and then increasing slowly over the time frame of the model without ever returning to 1. That is, the relative treatment effect increases until ~12 months after which it begins to</u>**

Page 76, last paragraph and Page 77, first paragraph "The manufacturer's base case assumption that the HR decreases over time suggests that not only is there a continued treatment effect, but that the treatment effect actually improves over time. This assumption is central to the period of extrapolation and is a key factor is explaining the large differences reported between the median OS gains (i.e. based on the observed data) and the mean OS gains (i.e. based on the observed data and the projected survival estimates including the extrapolation period) with aflibercept + FOLFIRI." **Page 109, second paragraph** "By fitting separate parametric curves to each treatment arm, it was assumed in the MS that the relative treatment effect on OS will continue to increase over the extrapolation period. That is, the OS curves continue to diverge during this period"	resulted in an increasing treatment effect during the first year and then a declining treatment effect thereafter. This resulted in converging OS curves during the extrapolation period"		**decrease (i.e. the survival curves begin to converge). However, the relative treatment estimate declines at a relatively slow rate over the remaining time horizon of the model and importantly suggests a continuing treatment effect on OS during the entire 15 year horizon.** **Pg 18, first paragraph** The ERG explored two assumptions other than the manufacturer's base case which assumed that the treatment effect of aflibercept **continued** for the full timeframe of the model. **Page 76, last paragraph** **The use of the independent parametric functions resulted in the hazard ratio decreasing from HR>1 to HR<1 over the first year and then increasing slowly over the time frame of the model without ever returning to 1. That is, the relative treatment effect increases until ~12 months after which it begins to decrease (i.e. the survival curves begin to converge). However, the relative treatment estimate declines at a relatively slow rate over the remaining time horizon of the model and importantly suggests a continuing treatment effect on OS during the entire 15 year horizon.** **Page 77, first paragraph** The assumption that the OS curves do not **fully converge** during the extrapolation period appears contrary to the PFS data from VELOUR that show that after approximately one year the PFS curves converge: there is no treatment difference after about 12 months (Figure 4.4). **Page 109, second paragraph** By fitting separate parametric curves to each treatment arm, it was assumed in the MS that the relative treatment effect on OS will **not fully converge** over the extrapolation period.

			Page 111, first paragraph Since the treatment effect may **not continue** for the entire duration of the model time horizon… ii) the difference in the treatment effect for OS declines over 12 or 18 months until there are no differences in the treatment arms (i.e. allowing for the curves to **fully** converge at an earlier time point).

ISSUE 6 ERG Scenario for OS curves converging at an earlier point

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
[Factual error] **Page 112, first paragraph** "The ERG also tested scenarios where the treatment effect on the aflibercept + FOLFIRI treatment would gradually decrease, similar to the convergence observed with the PFS data in the VELOUR trial. This scenario represents the possibility that the survival curves may converge at an earlier point. That is, it may be possible that patients on aflibercept + FOLFIRI actually face a higher risk of mortality during the unobserved period than patients on FOLFIRI alone."	Amendment of the text to "The ERG also tested scenarios where the treatment effect for aflibercept + FOLFIRI **is reversed (i.e. the hazard for aflibercept + FOLFIRI is higher than for FOLFIRI) in order to produce convergence of the OS curves** similar to the convergence observed with the PFS data in the VELOUR trial. This scenario represents the possibility that the survival curves may converge at an earlier point. That is, it may be possible that patients on aflibercept + FOLFIRI actually face a higher risk of mortality during the unobserved period than patients on FOLFIRI alone"	This analysis is not correctly described – the scenario does not explore a treatment effect for aflibercept + FOLFIRI that gradually decreases. As the ERG report recognises in the next sentence, it explores a scenario in which the treatment effect for aflibercept + FOLFIRI is reversed (i.e. patients on aflibercept + FOLFIRI actually face a higher risk of mortality during the unobserved period than patients on FOLFIRI alone). Furthermore, on page 111 of their report, the ERG incorrectly attributes this 'reverse-effect' scenario analysis to those proposed within the 2008 NICE Guide to the methods of technology appraisal. In that document, NICE proposes that "assumptions should include the limiting assumption of no further benefit as well as more optimistic assumptions." Specifically, in Section 5.7.3: "Alternative scenarios should be considered to compare the implications of different assumptions around extrapolation for the results. For example, for the duration of treatment effects scenarios might include when the treatment benefit in the extrapolated phase is: (i) nil; (ii) the same	This is not a factual inaccuracy. In this scenario the treatment effect decreases until there is convergence as in PFS. The following sentences give more detail of how this occurs, by assuming a higher mortality rate in the aflibercept + FOLFIRI arm.

		as during the treatment phase and continues at the same level; or (iii) diminishes in the long term." The proposal in bullet iii) is reflected in the MS base case.	

ISSUE 7 Extrapolation of OS based on small numbers of patients

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
[Factual error] **Page 16, first paragraph** "The extrapolation was based on a small numbers of patients still at risk past 30 months, making the extrapolation and the difference in OS uncertain"	Removal of the sentence	This statement is incorrect. The extrapolation of OS was not based on patients still at risk past 30 months; parametric survival functions were fitted to data for all patients and extrapolated beyond the end of trial follow-up. As is common in survival analysis, the earlier part of the curve, where there are observations, has a greater influence on the parameter estimates than the later part of the curve where there are fewer observations.	**ERG correction:** **Page 16, first paragraph** "The extrapolation was based on a population with **a small number of patients still at risk past 30 months, and there is considerable uncertainty to the extrapolation and difference in OS beyond the observed period.**"

ISSUE 8 Drug acquisition and administration costs

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
[Reviewer Error] **Page 16, last paragraph:** "The ERG also identified several errors and other areas of concern with the model. Firstly, the acquisition and administration costs were not related to the proportion of patients on treatment in the model" **Pages 107-108** "As discussed in section 5.2.8.1, the acquisition and administration	Deletion of "the acquisition and administration costs were not related to the proportion of patients on treatment in the model" Deletion of the analysis on page 107-108 and adjustment of the ERG base case analysis to remove this "correction". All statements related to "model programming errors" in the context of this issue should be deleted.	This is not an error in the manufacturer submission model. Drug and administration costs were accurately estimated making use of the mean duration of treatment observed in the trial, and accounting for dose reductions and dose delays. Specifically, we estimated the number of doses per 2-week cycle by multiplying the mean dose per administration by the number of administrations per 2-week cycle, and then multiplying by the mean number of two-week cycles of treatment. By taking into account the duration of treatment and the number of administrations per	The ERG acknowledges that this is not a model error, but rather a different interpretation of how the administration and acquisition costs of second-line treatment should be reflected in the model. The ERG have assumed that patients will receive the mean dose of aflibercept at each cycle they are in the secondline treatment state. Additionally, the manufacturer states in their justification for amendment that "the utility weight for the stable disease sub-states "on second line treatment"

costs can be reflected by using the mean dose in the VELOUR trial, to represent the dose reductions and delays, combined with the proportion of patients within the stable disease substates. The advantage of calculating the costs in this way is that the costs are calculated for the same proportion of patients to which the utilities have been applied."		model cycle, we fully capture both dose reductions and dose delays. Data specific to the population being analysed were applied (i.e. the ITT population or subgroup). The proportion of patients in the "on second-line treatment" health state was used in the model to explore the impact of possible differences in HRQL during treatment, rather than to estimate drug costs. The ERG "correction" reported in section 6.2 (page 107) appears not to account for dose delays, as the acquisition and administration costs are "reflected by using the mean dose in the VELOUR trial, to represent the dose reductions and delays, combined with the proportion of patients within the stable disease substates." The mean dose does not reflect dose delays; the mean number of administrations per 2-week model cycle is required to reflect dose delays. The ERG stated advantage of calculating the costs in this way is that the "costs are calculated for the same proportion of patients to which the utilities have been applied." There is no advantage to this approach over calculating costs directly from the trial data, since the latter already accounts for the proportion of patients on second-line treatment. Furthermore, in the base case analysis, the utility weight for the stable disease sub-states "on second line treatment" and "discontinued second-line treatment" are identical. We therefore assert that the approach in manufacturer submission model is not inaccurate as implied by the ERG and should not be described as an error. Furthermore, it appears that the ERG analysis is inaccurate as dose delays were not taken into account (using the mean	and "discontinued second-line treatment" are identical. ", which they are not as disutility from AEs is only applied to patients "on second line treatment". **ERG correction:** **Page 16, last paragraph:** "The ERG also identified other areas of concern with the model. Firstly, the acquisition and administration costs were not applied to all patients in the second line treatment health state of the model"

		number of administrations per 2-week model cycle). The ERG analysis is therefore expected to overestimate drug and administration costs.	

ISSUE 9 Very small numbers of patients at risk at the end of the OS curve

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
[Over simplification] **Page 17, third paragraph** "There were very small numbers of patients at risk in the OS analysis at the later time points, which reduces the reliability of the longer-term results" **Page 77, paragraph 3** "As described previously in section 4.2.3, there is high uncertainty around the OS survival estimates past 30 months due to the small number of patients at risk (n=10 and n=6 for aflibercept + FOLFIRI and FOLFIRI respectively) at 33 months and (n=1 and n=0 for aflibercept + FOLFIRI and FOLFIRI respectively) at 36 months."	Deletion of these two sentences	The small numbers of patients at risk at the later time points in the OS analysis does not reduce the reliability of the analysis used to estimate survival after trial follow-up because the primary reason for the small numbers of patients at risk at the later time points is the high proportion of events observed; i.e. most patients died during the trial follow-up. This in fact reduces the uncertainty in the OS analysis in comparison with datasets where there are a larger number of patients at risk at the later time points, as less extrapolation is required. Consider the situation where all patients in the trial had died during follow-up. In this case, there are no patients at risk at the end of trial follow-up, but there is little uncertainty in the OS analysis as the time of all death events is known. Conversely, if for example only 10% of patients had died, there are a large number of patients at risk but a much greater degree of extrapolation and therefore uncertainty. As discussed in Issue 5 the extrapolation of OS was not based solely on the later part of the Kaplan-Meier curve; parametric survival functions were fitted to data for all patients and extrapolated beyond the end of trial follow-up. The earlier part of the curve where there are more observations has a greater influence on the parameter estimates than the later part of the curve.	Table 4.3 in the ERG report, shows that there was a considerable number of patients censored during the trial period (209 for aflibercept + FOLFIRI vs. 154 for FOLFIRI). Even if a higher rate of events in the observed period reduces the uncertainty in the OS extrapolation, this is still influenced by those patients that have been lost to follow-up and for which it is unknown whether they have experienced the event (death). The small proportion of patients at risk compared to the censoring makes the observations at the end of the trial uncertain and makes extrapolation to 5 times the observed period subject to considerable uncertainty. **ERG correction:** **Page 17, third paragraph** There were very small numbers of patients at risk in the OS analysis at the later time points. **<u>The small portion of patients at risk compared to the number of patients censored</u>** increases the uncertainty surrounding the longer-term results" **Page 77, paragraph 3** "As described previously in section 4.2.3, there is high uncertainty around the OS survival estimates past 30

		Median FUP time: 22.3 months	months due **to the amount of censoring (209 for aflibercept + FOLFIRI and 154 for FOLFIRI) and** the small number of patients at risk (n=10 and n=6 for aflibercept + FOLFIRI and FOLFIRI respectively) at 33 months and (n=1 and n=0 for aflibercept + FOLFIRI and FOLFIRI respectively) at 36 months."

ISSUE 10 Analogy made between the convergence of PFS curves and potential convergence of OS curves

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
[Inconsistency] **Page 26, paragraph 2** The review concluded that the relationship varies considerably across different types of cancer, and even where strong evidence exists supporting a correlation between treatment effects, it is unclear how that should be converted into a quantified relationship between PFS and OS treatment effects in a cost effectiveness model. **Page 26, paragraph 3** The DSU report included five studies in mCRC, one of which investigated second-line therapy… **Page 26, paragraph 3** Giessen et al., Petrelli et al., and Chirila et al., referred to in corroboration of the DSU report are not adequately described and do not support the argument that presented that the VELOUR OS curves should mirror the convergence of the VELOUR PFS curves.	Acknowledgement in various parts of the document that the evidence-base exploring the relationship of PFS and OS in second-line mCRC is uncertain and that more research is required to adequately characterise this.	The report repeatedly relies upon an inferred relationship between PFS and OS and uses this as a basis to provide a rationale for the 'accelerated' earlier convergence of the OS curves in the base case model. However, the report provides contradictory insights on the relationship between PFS and OS as described in the literature, and goes on to state that actually post-progression survival may be more important in predicting OS than PFS in this setting. Furthermore, combined with an acknowledgement that there is limited evidence in second-line mCRC it would seem inconsistent and unreasonable to draw strong inferences on the likely time at which the extrapolated VELOUR OS curves begin to converge or the rate of convergence after trial follow-up from the shapes of the PFS curves. Particularly, as the evidence from one of the papers cited in support of this relationship, and arguably providing the most comprehensive evidence-base concludes…**that the findings are counterintuitive, based on heterogeneous data with too little information to adequately establish this relationship (Chirila et al.,)**	This is not a factual inaccuracy. The ERG report does not use the studies on the relationship between PfS and OS as suggested. The scenario run in which the OS curves reflect the PfS curve is presented by the ERG as one plausible scenario for the committee to consider.

		In VELOUR, the PFS curves start to converge after the median, and the curves coincided after approximately 80% of patients had relapsed. By contrast, the OS curves still appear to be diverging at the end of trial follow-up by which time by which time the estimated proportion of patients alive is <20%.Given the weakness of the evidence base for the relationship between PFS and OS and the robustness of the VELOUR trial it seems unreasonable to suggest that the time at which the OS curves begin to converge, or the rate of convergence, can be predicted from observations of PFS.	

ISSUE 11 Suitability of the median as an estimate of the Overall survival benefit

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
[Factual error / Inconsistency] **Page 40 paragraph 3 and page 50 paragraph 1** "The survival curves do not suggest that median survival is inappropriate to use in the economic analysis".	These statements should be removed.	This is an unreasonable statement given the nature of the VELOUR trial evidence and the fact that all sensitivity analysis produces an incremental mean OS benefit in excess of 3 months [Table 4.6 in the ERG report]. Furthermore, the restricted mean OS based on the KM curves represents a [redacted]% increase on the median value and the piece-wise calculation of the Hazard ratios indicates the median is an underestimate of the survival benefit. A recent paper by Davies et al.(2012) explored – using evidence from past NICE appraisals - the implications of estimating the mean OS of new cancer therapies. Importantly they comment …"Neither the observed medians nor restricted means have a relationship to the sum of all patients' survival times." Davies et al. Health Outcomes Research in Medicine (2012) 3, e25-e36	The ERG agree to this amendment.

ISSUE 12 Impact of using different distributions on mean OS

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
[Factual error] **Page 40 paragraph 6 and Page 47 paragraph 2:** "The mean OS benefit estimate varies considerably (from 3 months to [redacted] months) depending on which distribution is used, which indicates that the mean OS	This should be reworded to remove the final clause ("which indicates that the mean OS results presented may not be reliable".	This statement is incorrect. The variation in mean OS is due to the fact that some of the standard statistical distributions evaluated are a poor fit to the data. This is the reason for analysing statistical fit according to AIC and BIC and visual inspection	ERG correction **Page 40 paragraph 6 and Page 47 paragraph 2:** The mean OS benefit estimate varies considerably (from 3 months to [redacted] months) depending on which distribution is used, which indicates that the mean OS results presented are not

results presented may not be reliable".			robust to the choice of distribution and there is a high degree of uncertainty surrounding the mean OS results.

ISSUE 13 Critique of the Ruff et al piecewise analysis

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
[Factual error] **Page 45 paragraph 2** " The time periods used to calculate hazard ratios were unequal as all time points beyond 18 months were combined, making the HR unreliable for this time point".	The clause "making the HR unreliable for this time point" should be removed	This statement is incorrect - there is no reason per se why unequal time intervals should result in unreliable results. Indeed, as highlighted in our response to clarification questions, the reason for combining time points beyond 18 months was due to the small patient numbers at later time points.	The ERG accept this correction. This has been amended in the erratum.

ISSUE 14 Convergence of OS curves

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
[Inconsistency] **Page 46 paragraph 2** "Clinical experience with treatments for mCRC would suggest that extrapolation of the curves beyond the available data should reasonably assume their convergence at a later time point".	This statement should be removed.	This statement is somewhat of a non-sequitur. Clearly, at some point, all OS curves converge. This is the case both when a proportional hazards assumption is used and when treatment arms are modelled separately. This sentence is confusing and not relevant to the discussion of the approach taken (proportional hazards versus treatment arms separately).	This is not a factual inaccuracy. This is important since the mean estimates of OS from the manufacturer assume that the OS never fully converge during the time horizon of the model.

ISSUE 15 Cost-effectiveness data in the literature

Description of problem	Description of proposed amendment	Justification for amendment	ERG response

[Incomplete information] Page 66 paragraph 2 This paragraph describes the cost-effectiveness analysis of aflibercept versus bevacizumab and reports the results and mentions that the price of aflibercept is higher than that of bevacizumab	For completeness sake this should include that the aflibercept price in the abstract being described is substantially higher than either the current UK list price of aflibercept or the PAS price included in this submission	Given that the results of this abstract and a comment on the aflibercept price are included, it is reasonable to include a comment on the aflibercept price in the abstract versus the price of aflibercept in the current evaluation	As the ERG stated in the report, there is a lack of detail in the abstract. Furthermore, despite the price of aflibercept being mentioned in the abstract has higher than that of bevacizumab and a cost-effectiveness driver, the values used in economic evaluation were not disclosed. Therefore, the ERG could not verify whether the price of aflibercept in the abstract differed from that in the current evaluation.

ISSUE 16 Model population

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
[Factual error] **Page 67 paragraph 2** "The model considers aflibercept + FOLFIRI versus FOLFIRI alone as second-line treatment ofmCRC after an oxaliplatin-containing regimen. This is only a part of the overall licensed population which includes all patients who have previously used oxaliplatin and does not limit aflibercept use to second-line."	This should be removed.	The main inputs in the model come from the VELOUR ITT population and are therefore in line with the licensed indication as they are based on both patients receiving second-line treatment and patients who received oxaliplatin in the adjuvant setting only (approximately 10% of the VELOUR trial population). Therefore, we consider that the model is representative of the likely cost-effectiveness of aflibercept within the licensed population.	The license does not limit the use of aflibercept to second line treatment. The model does not consider aflibercept + FOLFIRI as first line for metastatic disease. This could occur if oxaliplatin has been used to treat less severe disease.

ISSUE 17 Search strategy – use of MeSH terms

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
[Factual error] **Page 29 – third paragraph** "Further, the MeSH terms used ('Colorectal Neoplasms', 'Colonic Neoplasms' and 'Rectal	Please remove this sentence altogether.	This is incorrect; the MeSH terms were exploded. Non-exploded MeSHh terms are signified in the search strategy by "[Mesh:NoExp]". Such terms were not used in the search strategy.	The ERG accept this correction. This has been amended in the erratum.

Neoplasms') were not exploded, and so additional MeSH terms found further down the MeSH hierarchy would not have been searched for."			

ISSUE 18 Inclusion of QoL data within the submission

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
[Factual error] **Page 17 paragraph 3 and Page 20 paragraph 4:** "There were no HRQoL data presented in the clinical section of the MS".	We suggest changing to "there were no HRQoL data collected in VELOUR"	While it is correct to say that no QoL data were collected in the VELOUR trial, baseline data from the ASQoP trial are presented in the clinical section, and information from an observational utility study is presented in the economic section).	**ERG correction:** **Page 17 paragraph 3 and Page 20 paragraph 4:** Substitute "There were no HRQoL data presented in the clinical section of the MS" with "**There were no HRQoL data collected in the VELOUR trial**"..

ISSUE 19 Report of an ongoing study that is not reporting within 12 months

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
[Relevance?] **Page 32 – second bullet point under section 4.1.7.** "AFLAME is a double-blind placebo-controlled RCT of aflibercept versus placebo with FOLFIRI in Asian patients with metastatic colorectal cancer previously treated with an oxaliplatin chemotherapy. The estimated completion date is January 2016."	Suggest removing mention of this from the report.	This study is not reporting in the next 12 months so does not need to be included here.	This is not a factual inaccuracy.

ISSUE 20 Description of subgroups

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
[Factual error] **Page 35 paragraph 2**: "The MS contains a post-hoc subgroup analysis, excluding patients who received oxaliplatin as adjuvant therapy".	This should be amended to say a "post-hoc subgroup which excluded patients who progressed on or within 6 months of oxaliplatin-based adjuvant therapy".	This subgroup excludes patients who progressed on or within 6 months of oxaliplatin-based adjuvant therapy without receiving 1st line treatment- for example a patient who received oxaliplatin as adjuvant therapy, then later received first-line chemotherapy, would still be included within this post-hoc subgroup.	ERG correction **Page 35 paragraph 2** The MS contains a post-hoc subgroup analysis, excluding patients who **progressed on or within 6 months of oxaliplatin-based** adjuvant therapy

ISSUE 21 Comparison of AEs between treatment groups

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
[Factual error] **Page 54, last sentence:** "With the exception of nausea, vomiting, alopecia, constipation, and the anti-VEGF-associated events 'fistula from other than GI origin' and GI perforation, which were similar between groups, all other adverse events were more common in the aflibercept group"	Please change to include mention of higher rate of anaemia in placebo group compared with aflibercept group. We suggest: "With the exception of anaemia, which was more common in the placebo group; and nausea, vomiting, alopecia, constipation, and the anti-VEGF-associated events 'fistula from other than GI origin' and GI perforation, which were similar between groups, all other adverse events were more common in the aflibercept group;…"	The current text is incomplete as written and could mislead the AC that all AEs are more common with aflibercept, or similar between groups.	The ERG agree to this amendment. This has been incorporated into the erratum.

ISSUE 22 Estimation of Adverse Event Disutilities

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
[Incomplete information] **Page 86, second paragraph:** However, when applying AEs related disutility in the model, the manufacturer appeared to divide the utility by the number of cycles	Deletion of the last sentence.	After correcting for the duplication of adjustment, we then conducted a one-way sensitivity analysis varying the utility decrement +/- 30% of the base case value. In doing so, we found that the ICER increased to only 38,719 when the decrement for aflibercept+FOLFIRI was	**ERG correction:** **Page 86, second paragraph:** "Hence, the apparent robustness of the model results **to changes in the absolute disutility values for each type of AE may also be in part**

in each year. As the estimate of disutility already accounted for the duration of the AEs converted to biweekly cycles, there appears to be a duplication of the adjustment for cycle length. The removal of the adjustment for AEs duration increased the base case ICER from £36,294 to £37,834 per additional QALY. **Hence, the apparent robustness of the model results to disutility of AEs may also be in part attributed to this error**.		increased 30% and 37,943 when the decrement for FOLFIRI was reduced by 30%. As such, the last sentence seems inappropriate. The impact of adverse events on the ICER is negligible, and the duplication of adjustment for AE disutility is not attributable to this error.	**attributed to this error, Repeating the analysis that simultaneously doubled disutility estimates for each type of AE on the corrected model increased the base case ICER by 9.05%."**

ISSUE 23 Generalisability of resource use study

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
[Incomplete information] **Page 88 paragraph 3** This paragraph questions whether the study population in the RU study was comparable to the patient population in UK routine clinical practice.	The text should be amended to reflect the design of the study and the inclusion criteria of patients included in this real-world study to allow the reader to interpret the relevance of the population.	The RU study was designed to sample records of patients within the NHS who would most likely be considered eligible for treatment with aflibercept. Whilst some parameters were not routinely recorded and therefore are not available for assessment, the nature of the inclusion criteria and the fact that the study centres are relevant UK centres should reassure the reader of the relevance of these data.	This is not a factual inaccuracy.

ISSUE 24 "Stop and drop" analysis

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
[Relevance to the decision] **Page 109-110** "The ERG first explored the sensitivity of the cost-effectiveness results to the time horizon of the model…….. a more appropriate basis	Deletion of the "stop and drop" analysis	The ERG report concedes that "this approach does not provide an appropriate basis for informing decision making" and that "this would likely provide an underestimate of the overall survival gains and result in an overly conservative estimate of the ICER."	This is not a factual inaccuracy.

for informing decision making would be to consider alternative scenarios concerning the treatment effect itself (or the shape of the survival functions) whilst maintaining a lifetime time horizon”		This analysis is of no value and could be misleading to the Appraisal Committee.	

ISSUE 25 Mean and median estimates from the KOL survey

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
[Clarification] **Page 119, paragraph 2** "However, the median resource use for community and personal and social care estimated from the KOL survey was applied to the model in the base case, instead of the mean. Mean values are most commonly used in cost-effectiveness analysis, as its aim is to estimate the expected values of costs and benefits. The use of medians is not common practice, and may underestimate expected costs. The use of the median was justified by the manufacturer based on the skew of the data."	Replace main document ERG base-case with the results provided in Appendix 1, based on the median estimates from the KOL survey, given that as the ERG acknowledged the results were: "considerably skewed".	We might expect the resources associated with treatment switches, the management of chemotherapy-related adverse events, and the lengths of stay of hospitalisations to be very much in the direct experience of the hospital-based oncologists who completed the survey. The elements of care subject to skew however, were the resource-use in the community setting, and consumption of personal social care resources, which perhaps in hindsight, were not in the day-to-day experience of these hospital-based experts, and may have been better estimated by frontline Community-based staff. That said, rather than dismiss this information, we considered the most appropriate way to handle this concern was to use the median values in the base case – we consider it is very unlikely that the differences between the extreme valuations of one or two experts really reflects the real distribution of resource used in these settings – and to test the effect of accommodating the more extreme values as a mean estimate within a sensitivity analysis. It is true that the effect of using the alternative data is a significant one on the ICER, but this is as a direct result of the skewed data increasing the cost of palliative care which disproportionately affects the aflibercept-treated patients as these patients enjoy greater survivorship within the model. Using the mean, rather than the median values, results in an increase of only ~£[redacted] in the incremental costs for stable disease. For progressive disease however, total costs in the aflibercept arm are increased by ~£[redacted] and by ~£[redacted] in the FOLFIRI arm, increasing the incremental costs for progressive disease by ~£[redacted]. As costs are applied on a per cycle basis costs are	This is not a factual inaccuracy.

		increased more in the aflibercept arm due to the longer post-progression survival	

ISSUE 26 Utility values – comparison with UK norms

[Factual error] **Page 83, paragraph 1** "The manufacturer argued that the sample of society from which the UK norms were derived included a significant proportion of individuals reporting disability (higher than in other national surveys). However, the ERG examined utility estimates from other national surveys, and found similar values to those reported by Kind et al., 1999 in the Health Survey for England."	This section of the report should be reworked to examine the assessment presented in the MS appropriately and to reflect directly the issues raised – rather than dismiss them by reference to other studies without further consideration or explanation.	The MS 'did not argue', but rather reported the findings of Kind et al. in 1998, 1999 – which using external survey data indicated that the population sampled (and since referenced as the 'normal' population) displayed greater levels of disability than other contemporaneous survey sources. Furthermore, the ERG report makes reference to a paper exploring utility values in patients with and without cardiovascular disease to corroborate the findings in Kind et al. This however misses the point made in our submission, which was not incidentally a criticism of the Kind et al research, but rather its use or indeed abuse in decision-making. Indeed, reference to the Sheffield paper ironically strengthens the evidence in support of our commentary. The principal issue is that the 'norm' data are drawn from a pool of respondents who cannot be considered 'disease-free'. Indeed, the 'average' valuation for each age group conceals a significant variation in individual respondent values – and does not, as many have perhaps assumed, represent the otherwise 'healthy individuals' at each age. Indeed, looking at the variation, for each age, in the Sheffield data highlights the significant range of values associated with cohorts characterised by CV status, in terms of CV events, both recent and in previous years, and those with	**ERG correction** **Page 83, paragraph 1** "The manufacturer **reported** that the sample of society from which the UK norms were derived included a significant proportion of individuals reporting disability (higher than in other national surveys). However, the ERG examined utility estimates from other national surveys, and found similar values to those reported by Kind et al., 1999 in the Health Survey for England."

		histories of single verses multiple events. What this means is that, by judging the actual, measured EQ5D scores from fit second-line mCRC patients, against population norms that are assumed to reflect healthy individuals, but in fact include significant variations due to other comorbidities, inappropriately prejudices opinion against the mCRC patient data. For example, in the Sheffield paper, it appears from the charts that respondents with no history of CVD, experience utility values in excess of 0.7 well into their mid-80s, but it is the presence of a past history of CVD events that 'dilutes' this average valuation. The MS submission was drawing attention to the apparently often overlooked fact that EQ5D values found in mCRC and in other cancer studies, whilst relatively high when considering the age and health status of the patients compared to the 'norm' values, but that this is nonetheless reasonable given the fact that for patients to be suitable for second-line chemotherapy, they must be fit enough to receive treatment, meaning they typically have a good performance status and reasonable HRQoL and importantly few comorbidities. It is therefore wrong, and incorrect to assert that a patients' directly reported EQ5D health-state, valued by societal preference weights, is in some way dubious or false, simply because there is a limited understanding or perhaps a misunderstanding of what the population 'norms' actually reflect.	

ISSUE 27 Subgroup base-case estimates

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
[Factual error] **Page 98 paragraph 2** "The assumptions for the subgroup specific base cases were not described in the MS".	This statement should be removed.	Appendix 14 details the parameterisation of OS and PFS for the subgroups, and also lists other parameters which differed between the subgroups (presented in Appendix 15).	**ERG correction** **Page 98 paragraph 2** "The assumptions for the subgroup specific base **cases were described in Appendix 14**."

ISSUE 28 Removal or adjustment of additional analyses that accommodate back-calculation of a confidential PAS

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
[Confidentiality] **Page 105, Figure 5.4 and associated text** Inclusion of sensitivity analysis related to drug acquisition costs	[redacted]	[redacted]	**Page 105, Figure 5.4 and associated text** – should be marked as CIC.

ISSUE 29 Assumptions around utility in the progressive disease state

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
[Clarification] **Page 83, paragraph 4** "The assumption that utility in this state is independent of time spent in the state is clinically implausible as patients are likely to experience deterioration in HRQoL throughout time as disease continues to progress."	This sentence should be revised as: "The assumption that utility in this state is independent of time spent in the state (with the exception of the end-of-life period) is, though clinically plausible, uncertain."	According to clinical experts consulted during the model-development advisory boards, we understand that the opposite proposal to that presented in the report is considered clinically plausible, in that patients' HRQoL does not continuously worsen in the progressive disease-state, not least because of palliative care management or the use of subsequent lines of treatment, but only towards the end of life do patients experience a significant worsening. The report challenges the clinical	This is not a factual inaccuracy.

		plausibility of this assumption, but offers no evidence to support this opinion and should acknowledge the current status of the evidence.	