15th July 2013

Meindert Boysen
Programme Director Technology Appraisals
Centre for Health Technology Evaluation
National Institute for Health and Care Excellence
Level 1A, City Tower
Piccadilly Plaza
Manchester
M1 4BT

Dear Meindert,

Re: Appraisal Consultation Document – aflibercept (Zaltrap) for the treatment of metastatic colorectal cancer

We are very disappointed by the Committee's initial decision not to recommend aflibercept within its licensed indication. We welcome the opportunity to address the Committee's concerns in this response document.

The Committee recognised that aflibercept is a novel agent with a significant treatment benefit on the three main clinical endpoints evaluated in the Phase III trial – overall survival, progression-free survival, and response rate. The Committee also acknowledge in their consultation document the feedback from clinical specialists and patients' representatives that current treatment options are limited and that it is important to make further treatment options available for patients.

We understand the Committee did not accept aflibercept on the basis of cost-effectiveness. The Committee accepted a number of revisions made by the ERG to our economic evaluation and based on this considered the most 'plausible ICER' fell within the range of £62K-£66K per QALY. The Committee also considered aflibercept did not fulfil the criteria for consideration as an end-of-life medicine under the supplementary advice from NICE.

As a treatment that has been demonstrated to extend survival in metastatic cancer which has progressed after first-line chemotherapy, we consider that aflibercept should be appraised under NICE's supplementary end-of-life criteria. We outline in our response below the evidence that supports why aflibercept meets all three criteria. We also address a number of concerns raised by the Committee on our economic model. We have provided a revised base-case economic model that incorporates a number of adjustments to address the Committee's concerns.

Together with a revised patient access scheme discount, we consider that aflibercept in combination with FOLFIRI is a cost-effective treatment option for metastatic colorectal cancer that has progressed after an oxaliplatin-containing regimen. We ask the Committee to reconsider its decision and approve aflibercept for use within the NHS in this area of high unmet medical need.

We have sought to address the Committee's main concerns in the following sections of this letter; we also provide a tabulated response to address other comments we have on the consultation document. In an Appendix we describe a revised base-case model and results.

Please do not hesitate to contact us if you require any further information.

Yours sincerely,

Charlie Nicholls
Head of Health Outcomes

1. *Revised Patient Access Scheme*

We will revise our patient access scheme discount. The revised cost of £[redacted] per vial (100 mg) and £[redacted] per vial (200 mg) represents a substantial discount of approximately [redacted] off the published list price.

The new PAS discount is included within our revised modelling analyses.

2. *Aflibercept should be evaluated under the provisions of the Supplementary advice to the Appraisal Committees on life-extending, end of life treatments*

The Committee noted the mean and median estimates of survival in the placebo group in VELOUR, and with the ERG's preferred estimate from the literature, concluded that patients receiving standard NHS treatment would have an expected survival of less than 24 months; therefore accepting that the first *end-of-life* criterion was met.

With regards to the second *end-of-life* criterion – that the treatment offers an extension to life of at least 3 months compared to standard treatment – we believe that the Committee is mistaken in referencing its decision to the median overall survival data "in the absence of a robust estimate of mean overall survival" since both they and the ERG recognise the median is clearly an underestimate of the true benefit. This is illustrated by the fact that even the restricted mean OS is 33% higher than the median and, as acknowledged by the ERG, is also likely to be an underestimate given that 17.2% of patients were alive in the aflibercept arm at the end of trial follow-up.

Furthermore, the Committee's rejection of aflibercept on the second *end-of-life* criterion is inconsistent with its conclusions on the most plausible ICER. The Committee acknowledge that the ERG's analysis which assumed no treatment effect in the follow-up period provides an 'acceptable compromise' and the 'most plausible assumption for extrapolating overall survival'. This preferred extrapolation approach itself produces a mean incremental survival gain of between 3.4 and 3.7 months depending on when the treatment effect is adjusted (e.g. 36m or 30m respectively; Table 1).

Table 1: Mean OS obtained assuming no continued treatment effect after 30 or 36 months

Extrapolation Function	Mean OS (months)
Log-logistic, with HR = 1 after 30 months	3.4
Log-logistic, with HR = 1 after 36 months	3.7

While it is reasonable to expect variation in the values derived from different extrapolation approaches – due to the different mathematical properties of the modelled functions and different modelling time horizons – we note that all the parametric functions assessed produce around three months additional survival, as shown in Table 2.

Table 2: Mean OS obtained with different parametric functions

	Incremental survival (months)		
Function	**5 year Time horizon**	**10 year Time horizon**	**15 year Time horizon**
Log-logistic	3.0	4.2	4.7
Weibull	2.9	3.0	3.0
Log-normal	3.4	4.8	5.3
Exponential	3.0	3.6	3.6
Gompertz	3.0	3.1	3.1

Given the consistent demonstration of at least 3 months survival gain, we request that the Committee afford aflibercept the benefit of the doubt as was the case for panitumumab for 2nd line

mCRC in TA242[1]. In considering whether panitumumab met the second end-of-life criterion, the Committee concluded that after adjusting for cross-over the mean overall survival gain of panitumumab was estimated to be between 2.7 and 3.2 months. The Committee also took into account – in spite of advice from the clinical experts that gains in progression-free survival cannot reliably translate to gains in overall survival – that panitumumab had demonstrated a progression-free survival gain similar to cetuximab (5 weeks and 7.8 weeks respectively, in KRAS wild-type patients) and since cetuximab had demonstrated an improvement in survival of over 3 months, panitumumab's survival advantage had been robustly demonstrated. Aflibercept increases progression-free survival by 9.6 weeks.

With respect to the third and final *end-of-life* criterion – that the treatment is licensed or otherwise indicated, for small patient populations – we remind the Committee that the population eligible for treatment with aflibercept (Zaltrap) is small; approximately 2600 of the 4000 patients estimated to receive any second line treatment for mCRC.

The ACD however, casts doubt on the acceptance of this last criterion in making reference to Eylea, a different formulation of aflibercept that is licensed separately for wet acute macular degeneration. The two medicines are not interchangeable, and aflibercept (Zaltrap) was recognised by the Department of Health as a different medicine and permitted Sanofi to price aflibercept (Zaltrap) as a "New Active Substance".

We therefore request that the Committee disregard the other treatment when applying the third *end-of-life* criterion and accept that aflibercept (Zaltrap) meets the 'small populations' test.

3. Comments on areas of uncertainty in the economic evaluation identified by the ERG and Appraisal Committee.

The ERG performed a number of adjustments to the economic model to address areas of uncertainty which they considered important (see below). From these alternative analyses, the Committee were able to select a "most plausible" ICER, which ranged between £62K to £66K per QALY.

The adjustments (on which the "most plausible" ICER was based) included:

Correction of AE disutility calculations

Adjustment to treatment acquisition costs so they were applied for entire duration patients were in "second-line on treatment" state

Incorporation of ageing disutility based on a starting age of 70

Use of mean, rather than median estimates from the KOL resource-use survey

Incorporation of additional administration costs (£15 pharmacy preparation costs)

Utility of 0.6 for the Progressive Disease Health state

Assumption of no continued treatment effect (i.e. the application of a hazard ratio equal to 1) after the end of trial follow-up.

We provide a point-by-point response to these adjustments below. In Appendix B, we present a revised model base-case and results.

3.1. Correction of AE disutility calculations

We accept the correction to the calculation of AE-related disutilities proposed by the ERG and apply this in our revised base-case.

3.2. Use of accurate treatment acquisition costs

Paragraph 4.10 of the ACD states: “It [the Committee] further heard that the acquisition costs of second-line treatments in the model did not depend on the proportion of patients in each state, and that they were calculated outside the model. The Committee agreed that the costs and QALYs in the stable disease health state were not specific to the two sub-states (“on second-line treatment” and “post second line treatment”).”

We interpret from this statement that the Committee did not reject our approach to the calculation of treatment acquisition costs. We have therefore retained our original approach within our revised base-case, but have provided further justification for this below.

We consider that our original approach to calculating treatment costs is the most accurate way of modelling treatment acquisition costs and is more appropriate than the alternative presented by the ERG. Although the overall impact on the ICER is relatively small, the effect of adopting the ERG approach is to inflate the amount of medicine consumed compared to that which was actually used within the VELOUR trial, and thereby bias-upwards the cost of treatment – this has greatest impact on the aflibercept regimen as will be seen below.

In the VELOUR trial, as in normal clinical practice, patients could skip one or more of the individual components of their treatment regimen and/or experience dose delays during any cycle. Reflecting the reality of what treatments were used to achieve the efficacy outcomes observed in VELOUR is important in deriving an unbiased estimate of the cost-effectiveness of aflibercept.

In the VELOUR trial, the mean duration of treatment with the aflibercept plus FOLFIRI regimen was [redacted] cycles versus [redacted] cycles in the FOLFIRI plus placebo regimen. The fact that these durations are longer than the mean number of administrations of any one component is due to dose delays and dose skipping.

The average administrations derived from using the ERG’s alternative modelling approach departs from that observed in the VELOUR trial. Table 3 demonstrates that for all regiment components, the alternative approach overestimates numbers of administrations and the associated costs by 2-4%. While these appear to be small differences, they nevertheless increase the total treatment costs and this has more impact for the aflibercept plus FOLFIRI arm.

Table 3: Comparison of cost and cycle estimations with ERG calculations

		Number of cycles			Costs		
		Actual data from VELOUR	Estimated from time in 2nd-line health state	Difference	Costing based on actual data from VELOUR	ERG costs based on time in 2nd-line health state	Difference
Aflibercept plus FOLFIRI	Aflibercept	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	Irinotecan	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	Leucovorin	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	5-FU	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
FOLFIRI		[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	Irinotecan	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	Leucovorin	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	5-FU	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

*- based on original PAS price

As our original approach more accurately reflects the treatment administered in VELOUR trial we maintain that this is a better approach to modelling this complex interplay of regimen components and we therefore maintain this as part of our revised base-case.

3.3. Application of age-related disutility with a correction for the age of the starting cohort

In our revised base-case we have incorporated the age-related disutility as applied by the ERG. However, based on the available evidence and the feedback of experts, we consider the age of the population who are likely to receive aflibercept plus FOLFIRI in UK practice is similar to that observed in VELOUR; we therefore apply a starting age of 60 in our revised base-case.

The ERG considered the starting age should be 70, as they believe the VELOUR trial population is younger and fitter than patients with mCRC seen in UK clinical practice. The ERG report originally cited information on the CRC population as a whole in support of their conclusions on the generalisability of the VELOUR population; they subsequently retracted this, and their erratum document acknowledges their conclusions are based on the feedback of one clinical expert. The subpopulation of patients who will be suitable to receive and benefit from multiple lines of combination chemotherapy (and therefore to be eligible for aflibercept) is not the same as the mCRC or CRC population as a whole.

Our recollection of responses provided to Committee's questions by the clinical experts at the first Committee meeting were that the VELOUR population does indeed reflect those patients in UK clinical practice likely to receive treatment with aflibercept in combination with FOLFIRI. Both these experts and others that we have heard from, have reflected that age is not a selection criterion for treatment, however to be suitable for second-line combination chemotherapy patients must be fit and with relatively few comorbidities; such patients are therefore likely to be younger than the mCRC population as a whole.

We provide below empirical data collected from UK observational studies of patients with mCRC receiving second-line chemotherapy. These indicate that the average age of patients likely to receive aflibercept plus FOLFIRI in UK clinical practice is indeed consistent with the average age of patients in the VELOUR trial (Table 4).

We therefore question the accuracy of the references made to this discussion in the consultation document, and importantly the application of ageing disutilities to a modelled population with the starting age of 70, and use the starting age of 60 in our revised base-case.

Table 4: Age from UK real-world studies

Study	Methodology	Inclusion criteria	Mean age
Observational RU study (described previously in Sanofi submission)	Service evaluation in 4 NHS hospitals	All mCRC patients who had received an irinotecan-based chemotherapy regimen following prior treatment with an oxaliplatin-based regimen including patients relapsing within 6 months of completion of oxaliplatin adjuvant chemotherapy (i.e. all patients within the licensed indication for aflibercept) (commenced irinotecan from 1st Jan 2008)	63
Kantar health chart review[2]	Prospective study - physician documented every patient with mCRC seen over 2-3 weeks	All mCRC patients seen by included physicians	60 (active 2nd line therapy) 71 (supportive care only)
IMS OncoAnalyzer databse [3, 4]	Largest commercially available oncology patient record database	UK patients receiving 2nd line chemotherapy for mCRC (and not enrolled in clinical trials)	63; Median age category 61 – 65
mCRC utilities study (described previously in Sanofi submission)	Cross-sectional study in 2 key UK centres (and 3 Dutch centres)	On 2nd or subsequent lines of treatment or best supportive care , received prior oxaliplatin in 1st line and had Eastern Cooperative Oncology Group (ECOG) performance status scores of 0-2 at 2nd line initiation	63

3.4. Median values from survey data – why this was an appropriate approach to take

In our original submission we applied median values from a key opinion leader survey to estimate resource-use components associated with community and social care. The clinicians surveyed came from similar, large UK centres, and could be expected to be familiar with secondary care resource use. As commented in the consultation document, if the sample of clinicians were appropriately homogenous, then the distribution of the data collected would be largely uniform, and it would be more appropriate to use mean estimates than median. Whilst we agree with the Committee's opinion on this point, the evidence of significant variation (skew) in a number of the reported resource-use values is indicative of a lack of homogeneity in the survey respondents rather than a representation of the natural variation in the resource-use they are reporting on. It is reasonable to expect the experts' experience and knowledge of care in the community and hospice setting would be more variable as they would be less familiar with day-to-day activities in these settings.

To demonstrate this, we examine the responses to estimate of the number of visits (per two week period) from a palliative care team to support patients in the progressed disease state; the responses were:

0.6; 1; 1; 1.3; 2; 8.

The estimate of 8 is clearly an outlier in this series and its impact is significant – the ICER based on the mean (2.32) of these values is approximately £4000 higher than when using the median (1.17).

We considered it very unlikely that differences of this nature between one or two experts really reflects the real distribution of resources used in these settings and considered that the median value in this set of skewed data would provide a better and more consistent estimate of resource use for the costing of community and social care.

Our approach is further supported when we compare the results of using the median values with the estimates used in other mCRC appraisals. In TA242 the per-cycle management costs for the progressed disease state in the Assessment Group model were £1039 per month[5]. This is reasonably close to the £[redacted] per month we obtain using the median of survey responses. Using the mean of responses from the survey generates a substantially higher monthly cost of £1757.

Using the mean, rather than the median values, results in an increase of only ~£200 in the incremental costs for stable disease, whereas for progressive disease, the incremental costs are increased by ~£1000. As costs are applied on a per cycle basis, costs are increased more in the aflibercept arm due to the longer post-progression survival. Thus the impact is to penalise aflibercept for increasing survival.

Therefore, overall we consider that the median is still the most appropriate way to account for the skewed data in this particular survey, and generates progressive disease costs that are consistent with those in previous appraisals. We apply this in our base-case, but also explore in scenario analyses the impact of using the mean values from the survey data in general, but excluding the outlier for the parameter of the number of visits from the palliative care team (the parameter described above). We also provide an analysis using per cycle management costs based on TA242.

3.5. Administration costings to reflect increased cost of delivery

In our original submission we used HRG codes to estimate the overall costs to the NHS of chemotherapy administration. Healthcare Resource Groups are: "standard groupings of clinically similar treatments which use common levels of healthcare resource" and are considered the most relevant cost sources when considering the overall cost to the NHS (http://www.hscic.gov.uk/hrg).

We considered that the code SB14Z would be relevant to both FOLFIRI alone and FOLFIRI in combination with aflibercept; this is the most costly of the published HRG codes and may indeed overstate the actual costs of delivering the standard chemotherapy regimen, particularly after the first cycle.

The ERG and Committee considered additional costs may be incurred for the administration of aflibercept in addition to FOLFIRI. As noted by the ERG it is difficult to accurately estimate these additional costs. They estimated a cost of an additional £15 for pharmacy preparation time, based on that previously estimated for bevacizumab and cetuximab. This was estimated by the Assessment Group in TA242 based on expert input from a pharmacist and applied in their model[5]. This additional cost was incorporated into the analysis that generated the "most plausible ICERs" (£62K – £66K per QALY). The ERG also estimated a figure of £45 per cycle for the additional administration time related to aflibercept treatment. This was not included within the £62 – 66k ICER estimates, and the ACD reports that the Committee considered the ICER could be higher if these costs were considered.

The Committee also considered that the pharmacy preparation costs for aflibercept could be higher than the £15 estimated by the ERG. We do not consider this to be the case - the preparation costs for aflibercept will be similar to bevacizumab and cetuximab, which are both biologic drugs requiring similar preparation methods and times. There is no reason to suppose aflibercept would require higher preparation costs than these other biologics. We therefore consider the £15 cost applied in TA242 to be the most appropriate cost estimate available.

We acknowledge the ERG's and Committee's concerns that aflibercept may incur additional costs for pharmacy preparation and administration above those accounted for within the HRG code used in our original model, and address these within our revised base-case. We use the ERG's cost estimates and incorporate the addition of both an additional £15 pharmacy preparation cost and a cost of £45 for additional administration time (i.e. an additional £60 per aflibercept cycle). These were applied in line with the average number of cycles of aflibercept received in the model.

3.6. Why utilities for stable and progressive disease health states should be based on values reported directly from patients, as per the NICE reference case

The utility values in our submission come from an observational real-world study in patients with mCRC undergoing second-line treatment. These data were collected in the UK and Netherlands in 2012/13 and meet the exact requirements of the NICE reference case; i.e. EQ-5D values reported directly from relevant patients and valued using the UK value-set.

In considering the face validity of the utility values presented in our submission, the ERG and Committee made reference to population 'norm' data for the corresponding age-group indicating that both stable and progressive disease health state valuations appear too high. This, we suggest, is a mistaken comparison because it is prefaced on the notion that without their cancer, the individuals should report values reflecting the population 'norm' data, and that the presence of their cancer should impose a further burden upon them and naturally reduce this utility valuation. This is mistaken because the population 'norm' data represents an average value and includes patients with good and moderate health, along with those with poor health including significant comorbidities that would preclude them from consideration for second-line chemotherapy.

What this means is that, by judging the actual, measured EQ-5D scores from fit mCRC patients selected as suitable for second-line chemotherapy, against population norms that are assumed to reflect healthy individuals, but in fact include significant variations due to other major comorbidities, risks inappropriately prejudicing opinion against the real – directly measured – EQ-5D data from the relevant patient cohort.

We therefore suggest that the EQ-5D values reported from a directly relevant patient population should not be dismissed as unreasonably high based on an oversimplified benchmarking process with population norm data.

In considering the value attributed to progressive disease, the ERG has selected, and the Committee subsequently accepted, a value of 0.6 that does not appear to be based on data, but rather on a 'sense' of appropriateness. We would reiterate that the data from our utility study is measured (as per the NICE reference case) in real patients who had progressed following second-line chemotherapy. The mean time to progression was [redacted] months therefore these are not values from patients who had recently progressed and therefore should be considered reasonably representative of the progressive disease state as a whole.

A key consideration raised by the Committee is that progressive disease utility is likely to decline through the course of the progressive disease and therefore a higher value (e.g. 0.71) may not be applicable over the entire course of the progressive disease state. Our understanding from the clinical experts is that patients' HRQoL does not continuously worsen in the progressive disease-state but in most cases does so sharply only towards the end of life. To address this alongside our revised base case we have explored through a scenario analysis the application of a much lower utility for the end of life period. In this scenario we apply a decrement equivalent to applying a utility of 0.3 during the last two months of life. This is described in Appendix A.

3.6.1. Update to the aflibercept safety and utility study - ASQoP

The VELOUR study did not include QoL data but to remedy this we have collected EQ-5D data through a Phase IIIb trial (the ASQoP trial). In our submission we provided baseline data from an interim analysis of this study, which is collecting EQ-5D data from patients receiving aflibercept in combination with FOLFIRI (details of this study were provided in our original submission).

Here we present a second interim analysis. This analysis provides data from patients receiving aflibercept in combination with FOLFIRI following progression on an oxaliplatin-based regimen and is therefore highly relevant to the decision problem. Further baseline data have been added at this update for more recently recruited patients, and data for patients recruited earlier who have since received their first few cycles of treatment are now available. The earliest patients into the study have received further cycles but due to limited numbers, data for the later cycles are not presented here.

At the data cut-off date for this interim analysis, 149 patients were screened and 116 had received at least 1 cycle. The mean age was 61. Currently available EQ-5D results are shown in Table 5.

Table 5: EQ-5D data from ASQoP interim analysis (full Global interim population)

Visit	N	Mean EQ-5D (SD)
Baseline visit	67	0.77 (0.22)
Cycle 3	63	0.78 (0.24)
Change from baseline	63	0.01 (0.24)
Cycle 5	30	0.79 (0.18)
Change from baseline	30	-0.02 (0.25)

We incorporate the ASQoP data presented above in our revised base-case for the stable disease state, using a weighted average of cycle 3 and 5 (i.e. patients on treatment). This gives a value of 0.78, which is similar to the value obtained from our observational utility study ([redacted]) used for the stable disease state in our original model. As there are not yet data available from ASQoP for patients in the progressed disease state we use the same data as in our original submission for this (i.e. the value of 0.71 from the observational utility study).

3.7. Overall survival

The Committee and the ERG have reflected that the estimate of mean overall survival is a major driver of cost-effectiveness in the model. This is to be expected in an oncology model, and by following the Decision Support Unit's best practice guide for extrapolating survival data, we presented in our submission the results of multiple curve fit analyses and performed the recommended interpretive steps before settling on the base-case estimate for our economic evaluation. The variation in the predicted mean survival results seen with the fitting of multiple curves to the data is not unexpected – these different functions have different underlying mathematical formulae and therefore generate different results. This is the reason for testing different functions and analysing both statistical and visual fit to the raw data.

We recognise the Committee is concerned that the extrapolation of mean OS is however a source of uncertainty in the cost-effectiveness estimates. The NICE Methods Guide (2008, Section 5.7.3) states that "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 as during the treatment phase and continues at the same level; or (iii) diminishes in the long term".

The first scenario is the most pessimistic of these. This is represented by the analysis conducted by the ERG that assumes a hazard ratio of 1 after the end of trial follow-up (30 or 36 months), and is that presented in the analyses generating the Committee's "most plausible" ICER.

The second scenario is more positive than our original submitted base-case assumption and therefore we have not considered this further.

The third scenario is consistent with our base-case in which the HR gradually increases back to 1. However, we recognise the concerns of the Committee and the ERG that in this analysis the increase in the HR is slow and the HR does not reach 1 during the time horizon of the model.

We therefore explore a more conservative assumption in our revised base-case.

We consider that, rather than assuming the HR reverts to 1 after the end of the observed period, it is more clinically plausible to assume the HR will increase to 1 gradually over a short period of time. We have therefore performed an analysis in which the HR is gradually increased (tapered) back to 1 over a period of 12 months, following the end of trial follow-up. Further details of how this was applied in the model are provided in the Appendix A. We incorporate this in our revised base-case, starting the taper at 36 months. We have explored in scenario analyses the impact of starting this tapering at 30 or 24 months.

3.8 Impact of addressing the points raised on the economic model

In Appendix B we present a revised economic model base-case that addresses the points above and present deterministic and probabilistic results and scenario analyses. The base-case ICER in this was £42,242 (incorporating the revised PAS).

4 – Table of additional comments

Paragraph	Comment	Sanofi response
3.9	"The manufacturer stated that the subgroup of patients with liver metastases only was recognised as a relevant clinical subgroup for metastatic colorectal cancer in Bevacizumab in combination with oxaliplatin and either fluorouracil plus folinic acid or capecitabine for the treatment of metastatic colorectal cancer (NICE technology appraisal guidance 212)"	This should actually refer to TA176 – Cetuximab for the first-line treatment of metastatic colorectal cancer
3.16	"However, the ERG considered that patients in the trial were potentially fitter and younger than those seen in UK practice, and so patients in clinical practice may not achieve the level of benefit reported in trial." Also : "Between 2007 and 2009, around 72% of patients diagnosed with colorectal cancer in the UK were aged 65 years or over. By contrast, in the VELOUR trial, only 33.5% of the aflibercept group and 38.9% of the placebo group were people aged 65 years or over."	As discussed in section 3.3, this assumption is incorrect because a comparison is being made to the entire population of patients diagnosed with CRC. While age is not a selection criterion for treatment, patients who are eligible for second-line combination chemotherapy are on the whole likely to be younger, fitter patients. Therefore, the UK CRC population as a whole is not a suitable reference point for comparison to VELOUR.
3.17	" The ERG noted that the hazard ratios for overall survival provided by the manufacturer by 6-month periods had wide confidence intervals at the later time points of the VELOUR trial because by this time many patients were no longer alive, leaving few patients at risk of dying (around 5% at 30 months). The ERG stated that wide confidence intervals reflect imprecise estimates, and that interpreting hazard ratios towards the end of the trial is highly uncertain, particularly at 30 months and 36 months."	The intention of the piecewise analysis of the VELOUR OS curve reported in the abstract from Ruff et al was to demonstrate the continued divergence of the OS curves. The original analysis looked at time periods of 0-6 months, 6 – 12 months, 12 – 18 months and >18 months. We provided an analysis for the time period post-18 months split into 6 month time intervals at the request of the ERG; as recognised by the ERG these did have wide confidence intervals, because, as with any survival analysis curve, there are fewer patients left alive and fewer events at the tail of the curve.
3.19	"The ERG noted that the estimate of mean overall survival varied considerably depending on which parametric function the manufacturer used, making the manufacturer's estimate of the difference in mean overall survival (4.7 months) unreliable. "	The fact that different estimates were obtained for mean OS with different functions is to be expected – this is the reason for testing different functions and assessing goodness-of-fit by visual and statistical methods, which is what we did, in line with best practice methodology. Alternative parametric functions tested (Weibull, lognormal, exponential and Gompertz) are a poorer fit to the observed trial

		data. The fact that these more poorly fitting functions generate different estimates does not make the estimate obtained with the better fitting log-logistic function unreliable.
3.40	"The ERG stressed that extrapolating the overall survival curves beyond the trial period is highly uncertain given that no data were available for more than 3 years' follow-up, and particularly that the progression-free survival curves separated and then converged at around 1 year."	It is normal in oncology trials to not have complete survival data, and to extrapolate in order to provide a complete survival curve and thereby to calculate the mean OS. VELOUR is a large trial, with a median follow-up of 22.23 months which compares well with other second-line mCRC trials. With regards to the progression-free survival curves, as highlighted in our response to the ERG report, there is no evidence to demonstrate a clear relationship between PFS and OS and in the absence of evidence it is inappropriate to assume the OS curves will necessarily follow the same pattern as the PFS curves.
3.40	"Nor did the manufacturer explore that the overall survival curves for aflibercept plus FOLFIRI and FOLFIRI alone may converge over the extrapolation period (that is, the treatment effect of aflibercept plus FOLFIRI gradually decreases from the point at which the trial ends) similar to the convergence observed with progression-free survival (in this scenario the risk of death may be higher in the aflibercept plus FOLFIRI group during the extrapolation period than in the FOLFIRI alone group)."	It should be noted that this analysis is not best described as one where the treatment effect "gradually decreases"; rather the analysis where the OS curves converge constitutes a scenario where the treatment effect decreases and is then reversed, such that (as recognised in the following clause) the risk of death is increased with aflibercept.
3.41	"Regarding the utility estimates in the model, the ERG had concerns about the generalisability of the manufacturer's observational study because the study population appeared to be younger than UK patients, and the proportion of patients who had an ECOG PS of 2 was lower than that seen in UK clinical practice."	In section 3.3 of the main document we provide a number of data sources which show that the average age both in the observational study and in VELOUR are in line with the UK patient population that is represented in the decision problem. The ERG's assessment of average age was (as they recognised) based on clinical opinion and as highlighted previously, comparisons with the CRC population as a whole are misleading.
3.42	"The ERG was concerned that the utility estimates used in the model from the manufacturer's utility study, as well as those reported in the literature, were high when compared with values used in previous appraisals of metastatic colorectal cancer, or with general UK population norms"	As highlighted in section 3.6 of the main document, we consider that comparisons with UK population norms are misleading.
3.52	"The ERG also presented results based on the probabilistic sensitivity analysis for selected scenario analyses. These suggested that the	As discussed in section 4.1 of the document we have identified an error in the probabilistic sensitivity

	deterministic results (section 3.51) underestimated the ICER by £5000 to £14,000 per QALY gained”	analysis coding. This had a minimal effect in our original analyses however when drug acquisition costs are applied based on time in the second-line “on treatment” state as in the ERG analyses rather than costed outside the health states, the impact is much bigger, resulting in the difference observed by the ERG. We have corrected this error; the resultant deterministic and probabilistic ICERs are now a very close match.
4.4	“The Committee heard from clinical specialists that the disease and demographic characteristics seen in patients in the VELOUR trial were not the same as those treated in UK clinical practice, in which patients are older”	As discussed above and in section 3.3, we consider that the available evidence suggests that patients in VELOUR were not notably younger than those in UK clinical practice who are treated with second-line combination chemotherapy
4.12	“The Committee understood that, in the ERG’s second scenario, it may be possible that patients receiving aflibercept plus FOLFIRI have a higher risk of death than those receiving FOLFIRI alone (that is, the hazard ratio may be greater than 1).”	This statement is correct; indeed our approximation of the ERG’s analyses (which appear to be approximately linear during the convergence period) show that the HR must change over time during that period; we estimated that HRs of up to 3.34 (18 month convergence) and 3.51 (12 month convergence) were required, with a mean of 1.84 and 1.64 respectively.
4.13	“The Committee considered that the value of 0.60 used by the ERG better reflected patients whose disease had progressed.”	In section 3.6 we discuss why we consider this to be incorrect.
4.13	“ In addition, the Committee noted that many QALYs in the model accrued after 5 years of treatment but, because the life expectancy of most people with metastatic colorectal cancer is shorter than 2 years, the Committee did not accept that the modelling of QALYs was in line with clinical experience of treating metastatic colorectal cancer”	This sentence seems at odds with the view of the ERG and accepted by the Committee that a time horizon longer than 5 years is appropriate. It is common in oncology modelling for QALYs to be accrued in the tail of the OS curve, when a relatively small number of patients are alive. In the model, only [redacted]% (FOLFIRI) and [redacted]% (aflibercept plus FOLFIRI arm) of the QALYs are accrued after 5 years.
4.15	“The Committee agreed that, if the sample of clinicians was appropriately homogenous, and reflected similar practices, the distribution of the data collected from the survey would be largely uniform, and it would be more appropriate to use mean estimates than median”	In section 3.5 of the main response document we provide further justification for our original decision to use median estimates.
4.20	“ …no robust evidence had been presented to indicate that aflibercept plus FOLFIRI compared with FOLFIRI alone offered a 3-month survival gain and therefore aflibercept plus FOLFIRI did not meet	In section 2 of the main response document we provide further justification that there is robust evidence that the increase in

	this criterion”	mean OS for aflibercept is at least 3 months. Indeed, in the OS analyses conducted by the ERG and preferred by the Committee, the mean OS was well over 3 months (3.4 – 3.7).
4.21	“It understood that aflibercept holds a marketing authorisation for treatment of a much larger population with neovascular (wet) age-related macular degeneration, but that this was a different formulation of aflibercept marketed by another company.”	In section 2 we reiterate that the alternative formulation and indication of aflibercept is not relevant to this appraisal.

5. References

1) NICE - TA242: Cetuximab (monotherapy or combination chemotherapy), bevacizumab (in combination with non-oxaliplatin chemotherapy) and panitumumab (monotherapy) for the treatment of metastatic colorectal cancer after first-line chemotherapy (review of TA150 and part review of TA118). January 2012

2) Kantar Health Chart Review 2012 – data-on-file

3) Rose M, Iqbal SU, Naoshy S, Mehta J, Chau I. (2012). Current Treatment Patterns in First-Line (1L) and Second-Line (2L) Treatment for Metastatic Colorectal Cancer (mCRC) in Western Europe. ESMO 2012.

4) IMS OncoAnalyzer – data-on-file.

5) Hoyle M, Crathorne L, Peters J, Jones-Hughes T, Cooper C, Napier M, Tappenden P, Hyde C (2013). The clinical effectiveness and cost-effectiveness of cetuximab (mono- or combination chemotherapy), bevacizumab (combination with non-oxaliplatin chemotherapy) and panitumumab (monotherapy) for the treatment of metastatic colorectal cancer after first-line chemotherapy (review of technology appraisal No. 150 and part review of technology appraisal No. 118): a systematic review and economic model. Health Technology Assessment, 17 (14).

6) Ara R, Brazier J. Health related quality of life by age, gender and history of cardiovascular disease: results from the Health Survey for England. Sheffield: School of Health and Related Research (ScHARR), The University of Sheffield; 2012.

Evidence submission appendix

Appendix A – description of revised economic model base-case

Revised Patient Access Scheme

We will revise our patient access scheme discount. The revised cost of £[redacted] per vial (100 mg) and £[redacted] per vial (200 mg) represents a substantial discount of approximately [redacted] off the published list price.

A revised base case economic evaluation to address areas of uncertainty identified by the ERG and Appraisal Committee.

The ERG performed a number of adjustments to the economic model to address areas of uncertainty which they considered important (see below). From these alternative analyses, the Committee were able to select a "most plausible" ICER, which ranged between £62K to £66K per QALY.

The adjustments (on which the "most plausible" ICER was based) included:

Correction of AE disutility calculations

Adjustment to treatment acquisition costs so they were applied for entire duration patients were in "second-line on treatment" state

Incorporation of ageing disutility based on a starting age of 70

Use of mean, rather than median estimates from the KOL resource-use survey

Incorporation of additional administration costs (£15 pharmacy preparation costs)

Utility of 0.6 for the Progressive Disease Health state

Assumption of no continued treatment effect (i.e. the application of a hazard ratio equal to 1) after the end of trial follow-up.

We addressed these adjustments in our revised base-case as follows:

1) *Correction of AE disutility calculations*

We accept the correction to the calculation of AE-related disutilities proposed by the ERG and apply this in our revised base-case.

2) *Treatment acquisition costs*

We consider that our original approach to calculating treatment costs is the most accurate way of modelling treatment acquisition costs.

3) *Application of age-related disutility with a correction for the age of the starting cohort*

In our revised base-case we have incorporated the age-related disutility as applied by the ERG. However, based on the available evidence and the feedback of experts, we consider the age of the population who are likely to receive aflibercept plus FOLFIRI in UK practice is similar to that observed in VELOUR; we therefore apply a starting age of 60 in our revised base-case.

4) *Median values from survey data – why this was an appropriate approach to take*

In our original submission we applied median values from a key opinion leader survey to estimate resource-use components associated with community and social care. The clinicians surveyed came from similar, large UK centres, and could be expected to be familiar with secondary care resource use. As commented in the consultation document, if the sample of clinicians were appropriately homogenous, then the distribution of the data collected would be largely uniform, and it would be more appropriate to use mean estimates than median. Whilst we agree with the Committee's opinion on this point, the evidence of significant variation (skew) in a number of the reported resource-use values is indicative of a lack of homogeneity in the survey respondents rather than a representation of the natural variation in the resource-use they are reporting on. It is reasonable to expect the experts' experience and knowledge of care in the community and hospice setting would be more variable as they would be less familiar with day-to-day activities in these settings.

To demonstrate this, we examine the responses to estimate of the number of visits (per two week period) from a palliative care team to support patients in the progressed disease state; the responses were:

0.6; 1; 1; 1.3; 2; 8.

The estimate of 8 is clearly an outlier in this series and its impact is significant – the ICER based on the mean (2.32) of these values is approximately £4000 higher than when using the median (1.17).

We considered it very unlikely that differences of this nature between one or two experts really reflects the real distribution of resources used in these settings and considered that the median value in this set of skewed data would provide a better and more consistent estimate of resource use for the costing of community and social care.

Our approach is further supported when we compare the results of using the median values with the estimates used in other mCRC appraisals. In TA242 the per-cycle management costs for the progressed disease state in the Assessment Group model were £1039 per month[5]. This is reasonably close to the £[redacted] per month we obtain using the median of survey responses. Using the mean of responses from the survey generates a substantially higher monthly cost of £1757.

Using the mean, rather than the median values, results in an increase of only ~£200 in the incremental costs for stable disease, whereas for progressive disease, the incremental costs are increased by ~£1000. As costs are applied on a per cycle basis, costs are increased more in the aflibercept arm due to the longer post-progression survival. Thus the impact is to penalise aflibercept for increasing survival.

Therefore, overall we consider that the median is still the most appropriate way to account for the skewed data in this particular survey, and generates progressive disease costs that are consistent with those in previous appraisals.

We apply this in our base-case, but also explore in scenario analyses the impact of using the mean values from the survey data in general, but excluding the outlier for the parameter of the number of visits from the palliative care team (the parameter described above). We also provide an analysis using per cycle management costs based on TA242.

5) *Administration costings to reflect increased cost of delivery*

We acknowledge the ERG's and Committee's concerns that aflibercept may incur additional costs for pharmacy preparation and administration above those accounted for within the HRG code used in our original model, and address these within our revised base-case. We use the ERG's cost estimates and incorporate the addition of both an additional £15 pharmacy preparation cost and a cost of £45 for additional administration time (i.e. an additional £60 per aflibercept cycle). These were applied in line with the average number of cycles of aflibercept received in the model.

6) Utilities for stable and progressive disease health states

In considering the value attributed to progressive disease, the ERG has selected, and the Committee subsequently accepted, a value of 0.6 that does not appear to be based on data, but rather on a 'sense' of appropriateness. We would reiterate that the data from our utility study is measured (as per the NICE reference case) in real patients who had progressed following second-line chemotherapy. The mean time to progression was [redacted] months therefore these are not values from patients who had recently progressed and therefore should be considered reasonably representative of the progressive disease state as a whole.

A key consideration raised by the Committee is that progressive disease utility is likely to decline through the course of the progressive disease and therefore a higher value (e.g. 0.71) may not be applicable over the entire course of the progressive disease state. Our understanding from the clinical experts is that patients' HRQoL does not continuously worsen in the progressive disease-state but in most cases does so sharply only towards the end of life. To address this alongside our revised base case we have explored through a scenario analysis the application of a much lower utility for the end of life period. In this scenario we apply a decrement equivalent to applying a utility of 0.3 during the last two months of life.

There was insufficient time to add a new health state to the model to capture the end of life period; therefore, a simple approach was taken to perform this analysis. The mean time spent in the progressive disease health state was first calculated for each treatment cohort, from the model results. The last 2 months of this time was assumed to be associated with a lower utility weight (0.30); the utility weight during the rest of the time in the progressive disease state was assumed to be 0.71 (the base-case value for progressive disease in the submission model). Using the two time periods and two utility values, an adjusted utility weight was calculated for each treatment group and applied to all time spent in the progressive disease health state.

6.1.) Update to the aflibercept safety and utility study - ASQoP

The VELOUR study did not include QoL data but to remedy this we have collected EQ-5D data through a Phase IIIb trial (the ASQoP trial). In our submission we provided baseline data from an interim analysis of this study, which is collecting EQ-5D data from patients receiving aflibercept in combination with FOLFIRI (details of this study were provided in our original submission).

Here we present a second interim analysis. This analysis provides data from patients receiving aflibercept in combination with FOLFIRI following progression on an oxaliplatin-based regimen and is therefore highly relevant to the decision problem. Further baseline data have been added at this update for more recently recruited patients, and data for patients recruited earlier who have since received their first few cycles of treatment are now available. The earliest patients into the study have received further cycles but due to limited numbers, data for the later cycles are not presented here.

At the data cut-off date for this interim analysis, 149 patients were screened and 116 had received at least 1 cycle. The mean age was 61. Currently available EQ-5D results are shown in Table 5.

Table 1: EQ-5D data from ASQoP interim analysis (full Global interim population)

Visit	N	Mean EQ-5D (SD)
Baseline visit	67	0.77 (0.22)
Cycle 3	63	0.78 (0.24)
Change from baseline	63	0.01 (0.24)
Cycle 5	30	0.79 (0.18)
Change from baseline	30	-0.02 (0.25)

We incorporate the ASQoP data presented above in our revised base-case for the stable disease state, using a weighted average of cycle 3 and 5 (i.e. patients on treatment). This gives a value of 0.78, which is similar to the value obtained from our observational utility study ([redacted]) used for the stable disease state in our original model. As there are not yet data available from ASQoP for patients in the progressed disease state we use the same data as in our original submission for this (i.e. the value of 0.71 from the observational utility study).

7) Overall survival

We recognise the Committee is concerned that the extrapolation of mean OS is a source of uncertainty in the cost-effectiveness estimates.

We consider that, rather than assuming the HR reverts to 1 after the end of the observed period, it is more clinically plausible to assume the HR will increase to 1 gradually over a short period of time. We have therefore performed an analysis in which the HR is gradually increased (tapered) back to 1 over a period of 12 months, following the end of trial follow-up.

As with the ERG analyses, the individual OS functions were used during the trial follow-up period (36/ 30 months) and for the FOLFIRI arm, this function was continued to the end of the analysis. To explore a gradual diminution in the treatment effect over time, the hazard ratio for OS was assumed to taper to a value of 1.00 over a period of 12 months after the end of trial follow-up (30 or 36 months). A linear taper was assumed. The starting value of the hazard ratio at the beginning of the taper period was the hazard ratio observed at the end of trial follow-up. This was estimated from the individual OS functions as the ratio of the instantaneous hazard for aflibercept plus FOLFIRI versus FOLFIRI at the end of trial follow-up, each of which was calculated from the respective OS function.

We incorporate this in our revised base-case, starting the taper at 36 months. We have explored in scenario analyses the impact of starting this tapering at 30 or 24 months.

Appendix B - Revised base-case results

In this section we present results from a revised base-case to address the main concerns of the Committee. This revised base-case now includes:

- Incorporating the ERG's correction to AE-related disutility calculations
- Incorporating a disutility due to ageing (with the cohort age set to 60)
- Incorporating additional administration costs – an additional cost of £60 is applied per aflibercept cycle (including £15 pharmacy preparation time and £45 additional administration cost for aflibercept)
- Incorporation of ASQoP data for the Stable Disease utility in the model
- Tapering of the HR from the end of trial follow-up (36 months) over a 12 month period.

Results of revised modelling base-case

Results of the revised model base-case are shown in Table 7 and Table 8. These are shown as both probabilistic and deterministic ICERs to allow comparison. The new base-case ICER is £42,242/QALY.

We also draw attention here to the fact that, in their analyses, the ERG found a substantial discrepancy between deterministic and probabilistic results (estimated at £5000 – £14,000). This was due to a minor coding error in the original PSA; this had a minimal effect on the results for our original base-case but that had a much greater impact when treatment acquisition costs were applied according to duration in the second-line on-treatment state as in the ERG's analyses. We have corrected this error and now, importantly, the deterministic and probabilistic ICERs are very similar. We present both the probabilistic and deterministic ICERs for our revised base-case. However, due to time constraints, only deterministic ICERs are presented for the scenario analyses.

Table 2: Revised model base-case results - deterministic

	Total costs	Total LYGs	Total QALYs	Incremental costs	Incremental LYG	Incremental QALYs	ICER incremental (QALYs)
FOLFIRI	[redacted]	[redacted]	[redacted]	£8,500	0.289	0.201	£42,242
Aflibercept + FOLFIRI	[redacted]	[redacted]	[redacted]				

Table 3: Revised model base-case results - probabilistic

	Total costs	Total LYGs	Total QALYs	Incremental costs	Incremental LYG	Incremental QALYs	ICER incremental (QALYs)
FOLFIRI	[redacted]	[redacted]	[redacted]	£8,499	0.289	0.201	£42,197
Aflibercept + FOLFIRI	[redacted]	[redacted]	[redacted]				

The results of the PSA are presented on the cost-effectiveness plane in Figure 1 and as a cost-effectiveness acceptability curve in Figure 2. At a willingness-to-pay (WTP) of £50,000 per QALY the probability of cost-effectiveness was 72%.

Figure 1: PSA results presented on the cost-effectiveness plane

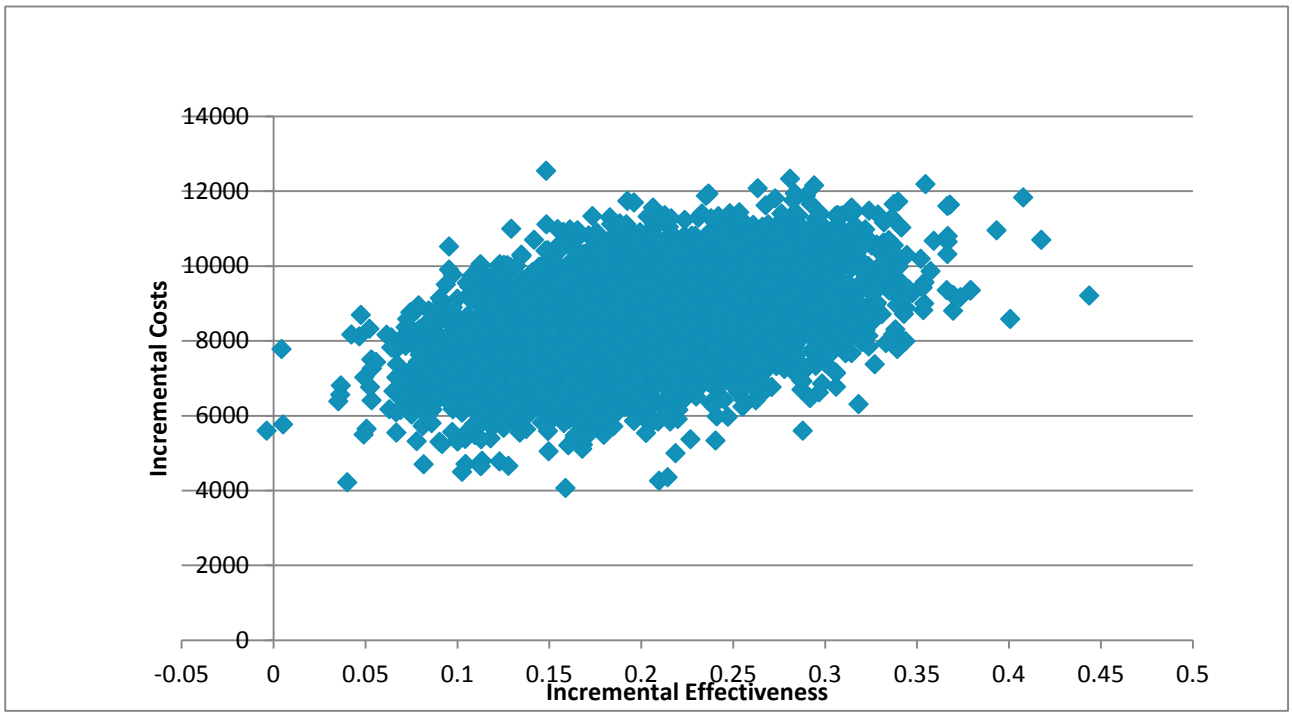
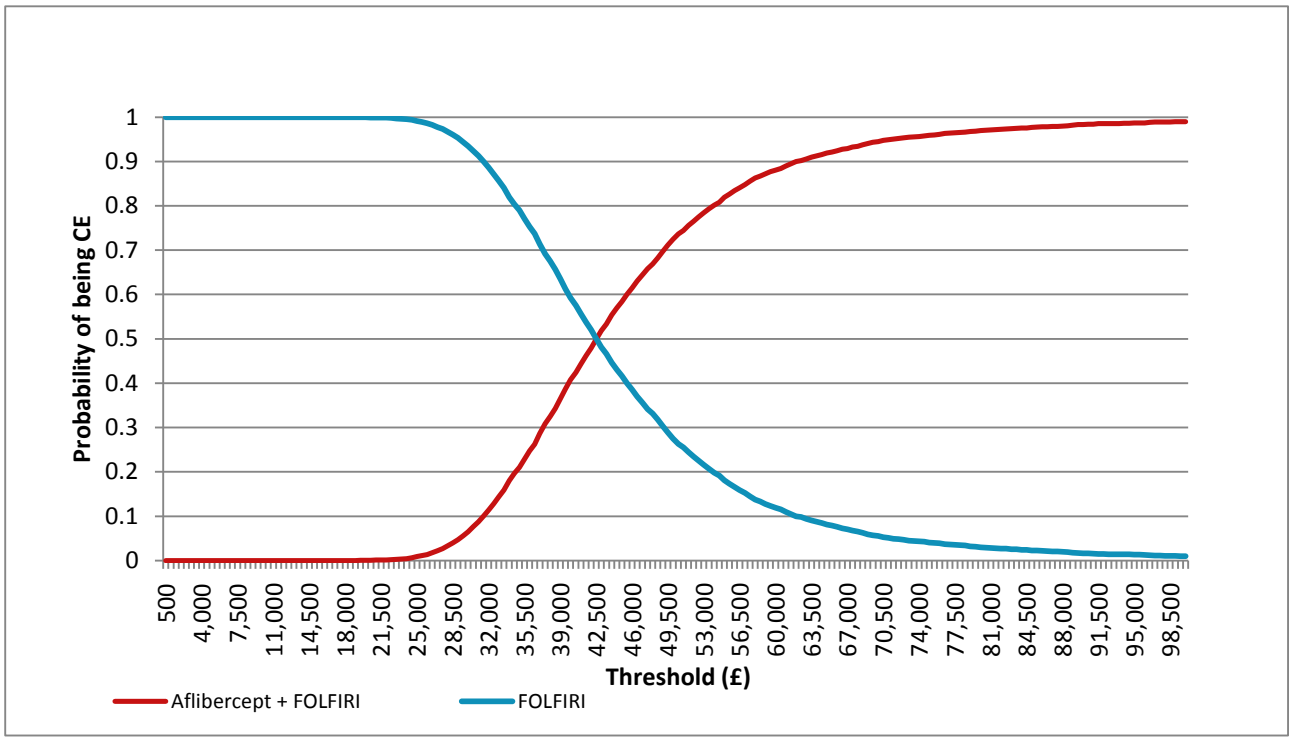


Figure 2: Cost-effectiveness acceptability curves



Results of the scenario analyses

Results of the scenario analyses are shown in Table 10.

Ageing disutility:

Incorporating a starting age of 65 had a minimal impact on the results.

KOL RU estimates:

As noted previously, using the mean rather than the median from the KOL survey has a substantial impact on results, increasing the ICER by ~£5K/QALY. The main driver of this appears to be the estimate of the "number of visits from palliative care team" – adjusting this to remove the outlier estimate and taking the mean of the remaining values and retaining the mean for all the other parameters gives an ICER much closer to the base-case ICER.

Incorporating the estimate for per cycle management costs from TA242 rather than from our own data gives an ICER extremely close to our base-case ICER when the median values are used.

Utility values:

Utility values from ASQoP data for stable disease were very similar to those obtained in the mCRC utilities study and as expected the resultant ICER was very similar with both datasets.

Adapting the model to assume a sharp drop in utility at the end of life had only a small impact on the ICER – this reflects the fact that all patients are expected to experience a rapid decline in the terminal stages of life and therefore cancel out in the analysis – the duration of this period is unlikely to differ between treatment arms.

Overall Survival extrapolation:

As would be expected, tapering the HR from an earlier time-point (i.e. 30 months instead of from 36 months) raised the ICER by £2K/QALY.

Tapering from 24 months, which we consider to be a pessimistic scenario given that this is within the observed follow-up period and the Kaplan-Meier curves clearly show continued divergence at this time-point, gives a new ICER of 47K/QALY.

Applying the assumption of no continued treatment effect at the end of trial follow-up, as in the ERG analyses, gives an ICER of £44K (36 months) and £47K/QALY (30 months).

Table 4: Scenario analysis results

Parameter	Base-case	Scenario	Incremental costs[a]	Incremental QALYs[a]	ICER
Base case results			£8,500	0.201	£42,242
Ageing disutility	Starting age 60	Starting age 65	£8,500	0.199	£42,740
KOL RU data	Median values	Adjustment to "number of visits from palliative care team" to remove outlier estimate; remaining values based on mean	£8,801	0.201	£43,740
		Per cycle management costs based on TA242	£8,548	0.201	£42,481
Utility values	ASQoP data used for SD; mCRC observational utilities study for PD	SD value from mCRC observational utilities study	£8,500	0.198	£42,943
		Decrement for end of life	£8,500	0.202	£42,002
OS	HR tapers to 1 over 12 month period – start at 36 months	HR tapers to 1 over 12 month period – start at 30 months	£8,419	0.191	£44,136
		HR tapers to 1 over 12 month period – start at 24 months	£8,302	0.176	£47,246
		HR = 1 after 36 months	£8,422	0.191	£44,039
		HR = 1 after 30 months	£8,307	0.176	£47,097

5. *References*

1) NICE - TA242: Cetuximab (monotherapy or combination chemotherapy), bevacizumab (in combination with non-oxaliplatin chemotherapy) and panitumumab (monotherapy) for the treatment of metastatic colorectal cancer after first-line chemotherapy (review of TA150 and part review of TA118). January 2012

2) Kantar Health Chart Review 2012 – data-on-file

3) Rose M, Iqbal SU, Naoshy S, Mehta J, Chau I. (2012). Current Treatment Patterns in First-Line (1L) and Second-Line (2L) Treatment for Metastatic Colorectal Cancer (mCRC) in Western Europe. ESMO 2012.

4) IMS OncoAnalyzer – data-on-file.

5) Hoyle M, Crathorne L, Peters J, Jones-Hughes T, Cooper C, Napier M, Tappenden P, Hyde C (2013). The clinical effectiveness and cost-effectiveness of cetuximab (mono- or combination chemotherapy), bevacizumab (combination with non-oxaliplatin chemotherapy) and panitumumab (monotherapy) for the treatment of metastatic colorectal cancer after first-line chemotherapy (review of technology appraisal No. 150 and part review of technology appraisal No. 118): a systematic review and economic model. Health Technology Assessment, 17 (14).

6) Ara R, Brazier J. Health related quality of life by age, gender and history of cardiovascular disease: results from the Health Survey for England. Sheffield: School of Health and Related Research (ScHARR), The University of Sheffield; 2012.