Evidence Review Group Report

Aflibercept in combination with irinotecan and fluorouracil-based therapy for the treatment of metastatic colorectal cancer which has progressed following prior oxaliplatin-based chemotherapy

Produced by CRD and CHE Technology Assessment Group, University of York, Heslington, York YO10 5DD

Authors Ros Wade, Research Fellow, CRD
Ana Duarte, Research Fellow, CHE
Mark Simmonds, Research Fellow, CRD
Rocio Rodriguez-Lopez, Information Specialist, CRD
Steven Duffy, Information Specialist, CRD
Eldon Spackman, Research Fellow, CHE
Nerys Woolacott, Senior Research Fellow, CRD

Correspondence to Nerys Woolacott, Senior Research Fellow, CRD, University of York YO10 5DD. Tel: (01904) 321074. Email: nerys.woolacott@york.ac.uk

Date completed 22 April 2013

Source of funding: This report was commissioned by the NIHR HTA Programme as project number 11/80/01.

Declared competing interests of the authors
None.

Acknowledgements
We would like to thank Dr Daniel Swinson (Consultant Medical Oncologist, Leeds Teaching Hospitals NHS Trust) for clinical advice throughout the project, and for providing comments on the draft ERG report. We would also like to thank Professor Stephen Palmer (Professor of Health Economics, CHE) for advice and for providing comments on the draft ERG report.

Rider on responsibility for report

The views expressed in this report are those of the authors and not necessarily those of the NIHR HTA Programme. Any errors are the responsibility of the authors.

This report should be referenced as follows:

Wade R, Duarte A, Simmonds M, Rodriguez-Lopez R, Duffy S, Spackman E, Woolacott N. Aflibercept in combination with irinotecan and fluorouracil-based therapy for the treatment of metastatic colorectal cancer which has progressed following prior oxaliplatin-based chemotherapy: A Single Technology Appraisal. CRD and CHE Technology Assessment Group, 2013.

Contributions of authors

Ros Wade and Nerys Woolacott wrote the clinical effectiveness sections of the report, with assistance from Mark Simmonds. Eldon Spackman managed and supervised the cost-effectiveness work and contributed to the writing of the cost effectiveness sections and the economic analyses. Ana Duarte wrote the cost effectiveness sections of the report and conducted the economic analyses. Rocio Rodriguez-Lopez and Steven Duffy wrote the sections on the search strategies.

Table of Contents

List of abbreviations 9
1. SUMMARY 12
1.1 Critique of the decision problem in the manufacturer's submission 12
1.2 Summary of clinical effectiveness evidence submitted by the manufacturer 12
1.3 Summary of the ERG's critique of clinical effectiveness evidence submitted 13
1.4 Summary of cost effectiveness evidence submitted by the manufacturer 14
1.5 Summary of the ERG's critique of cost effectiveness evidence submitted 15
1.6 ERG commentary on the robustness of evidence submitted by the manufacturer 17
1.6.1 Strengths 17
1.6.2 Weaknesses and areas of uncertainty 17
1.7 Summary of exploratory and sensitivity analyses undertaken by the ERG 18
1.8 Conclusions or key issues 18
2 BACKGROUND 20
2.1 Critique of manufacturer's description of underlying health problem 20
2.2 Critique of manufacturer's overview of current service provision 20
3 CRITIQUE OF MANUFACTURER'S DEFINITION OF DECISION PROBLEM 22
3.1 Population 22
3.2 Intervention 22
3.3 Comparators 23
3.4 Outcomes 25
3.5 Other relevant factors 27
4 CLINICAL EFFECTIVENESS 28
4.1 Critique of the methods of review(s) 28
4.1.1 Search strategy 28
4.1.2 Inclusion criteria 30
4.1.3 Data extraction 31
4.1.4 Quality assessment 31
4.1.5 Evidence synthesis 31
4.1.6 Conclusions from critique of systematic review methods 31
4.1.7 Ongoing studies 32
4.2 Critique of trials of the technology of interest, their analysis and interpretation (and any standard meta-analyses of these) 32

4.2.1 Trials included in the review 32
4.2.2 Summary of the quality of the included trials 43
4.2.3 Summary of the results of the included trials 44
4.3 Conclusions of the clinical effectiveness section 59
4.4 Other relevant evidence for consideration 60
4.4.1 Evidence for other anti-VEGF therapy (bevacizumab) 60
4.4.2 Additional evidence of aflibercept for the treatment of mCRC 62
5 COST EFFECTIVENESS 64
5.1 ERG comment on manufacturer's review of cost-effectiveness evidence 64
5.1.1 Searches 64
5.1.2 Inclusion/exclusion criteria used for study selection 65
5.1.3 Studies included and excluded in the cost effectiveness review 66
5.1.4 Conclusions of the systematic review 66
5.2 Summary and critique of manufacturer's submitted economic evaluation by the ERG…….....66
5.2.1 Model structure 67
5.2.2 The manufacturer's economic evaluation compared to the NICE reference case checklist.. 73
5.2.3 Population 73
5.2.4 Interventions and comparators 74
5.2.4.1 Appropriateness of comparators 74
5.2.5 Perspective, time horizon and discounting 75
5.2.6 Treatment effectiveness and extrapolation 75
5.2.6.1 PFS estimation 76
5.2.6.2 OS extrapolation 76
5.2.6.3 Treatment duration 77
5.2.7 Health related quality of life 79
5.2.7.1 Other sources of HRQoL data 80
5.2.7.2 Post-progression HRQoL 83
5.2.7.3 HRQoL adjustment for adverse events 84
5.2.7.4 Summary of Health Related Quality of Life 86
5.2.8 Resources and costs 87
5.2.8.1 Drug acquisition costs 89

5.2.8.2 Duration of second-line treatment 89
5.2.8.3 Drug administration costs 92
5.2.8.4 Key Opinion Leader Survey 93
5.2.8.5 Adverse events 94
5.2.8.6 Unit cost inconsistencies 97
5.2.9 Discounting 97
5.2.10 Cost effectiveness results 98
5.2.10.1 Base case 98
5.2.10.2 Sensitivity analyses 99
5.2.11 Model validation 104
5.3 Exploratory and sensitivity analyses undertaken by the ERG 105
5.4 Conclusions of the cost effectiveness section 105
6 IMPACT ON THE ICER OF ADDITIONAL CLINICAL AND ECONOMIC ANALYSES UNDERTAKEN BY THE ERG 107
6.1 Overview 107
6.2 ERG corrections and adjustments to the manufacturer's base case model 107
6.3 Additional ERG sensitivity analyses 109
6.3.1 Treatment effect duration 109
6.3.2 Post-progression utility 116
6.3.3 Additional administration costs for aflibercept 117
6.4 ERG alternative base case and OS scenarios 118
6.5 Subgroups 124
6.5.1 Liver metastases only subgroup 124
6.5.2 Excluding adjuvant chemotherapy subgroup 125
6.6 Conclusions 128
7 END OF LIFE 129
8 OVERALL CONCLUSIONS 131
8.1 Implications for research 131
References 132
Appendix 1 Cost effectiveness results for the ERG alternative scenarios using the median KOL survey resource use for health states 136
Appendix 2 Quality assessment using the Philips[36] economic modelling checklist 138

Table of Tables

Table 4.1 Design and methods of the VELOUR trial 33
Table 4.2 Baseline characteristics of patients in the VELOUR trial 36
Table 4.3 Overall survival (months), piecewise Cox model with pre-defined intervals 45
Table 4.4 AIC and BIC test results for overall survival functions 47
Table 4.5 Estimated mean overall survival at 15 years 48
Table 4.6 Estimated mean overall survival at different time horizons 49
Table 4.7 Response rates 54
Table 4.8 Most frequent adverse events (events (incidence ≥20% or ≥5% higher in the aflibercept arm), other anti-VEGF-associated events, and most frequent biologic abnormalities: safety population 56
Table 4.9 All further anti-cancer therapy 59
Table 4.10 Efficacy of other anti-VEGF therapy (bevacizumab) 62
Table 5.1 Summary of the manufacturer's economic evaluation, with justifications for key aspects and signposts to the relevant sections of the MS 69
Table 5.2 A comparison to the NICE reference case 73
Table 5.3 Summary of utility weights in the economic evaluation (model inputs emboldened) 80
Table 5.4 Selection of utility estimates from Best *et al.*, 2010 [45] 82
Table 5.5 Summary of utility estimates in TA212[43] 82
Table 5.6 Data used to estimate disutility due to adverse events 84
Table 5.7 Summary of QALY by health-state (Table B41, MS) 86
Table 5.8 Distribution of patients across further lines of treatment by model health state ..88
Table 5.9 List and model price of aflibercept 89
Table 5.10 Costs and doses of second-line treatment - adapted from the MS (Table B35)91
Table 5.11 Chemotherapy delivery unit costs (day case and regular day/night) [52] 93
Table 5.12 List of adverse events included in the model – adapted from the MS (Table B37)…… 95
Table 5.13 Impact on cost-effectiveness estimates of correcting AEs costing inconsistencies 96
Table 5.14 Base case cost-effectiveness results - adapted from the MS (Table B44) 98
Table 5.15 Base case cost-effectiveness results - adapted from the MS (Table B46) 99
Table 5.16 Results of the scenario analyses (from Table B45, of the MS) 101

Table 6.1 Utility decrement associated with ageing (unadjusted for cycle length) 108
Table 6.2 Cost-effectiveness results for different assumptions on treatment effect duration…….....115
Table 6.3 Utility estimates in Progressive Disease for mCRC 116
Table 6.4 Cost-effectiveness results for alternative assumption on PD utility 117
Table 6.5 Chemotherapy delivery unit costs (day case and regular day/night) [52] 117
Table 6.6 Cost-effectiveness results for alternative assumptions on administration costs .118
Table 6.7 Cost effectiveness results for the ERG alternative scenarios for the overall population, assuming mean resource use. 121
Table 6.8 Selected PSA cost effectiveness results for the ERG alternative scenarios overall population.....123
Table 6.9 Cost effectiveness results for the ERG revised scenarios for the Liver Metastases Only Subgroup 126
Table 6.10 Cost effectiveness results for the ERG revised scenarios for the Excluding Adjuvant Chemotherapy Subgroup 126

Table of Figures

Figure 4.1 Kaplan-Meier estimates of overall survival.....44

[redacted]

[redacted]…….....46

Figure 4.3 Overall survival by pre-specified subgroups.....51

Figure 4.4 Kaplan-Meier estimates of progression free survival.....52

Figure 5.1 Model structure (Figure B-10, of the MS).....68

Figure 5.2 Manufacturer's deterministic sensitivity analyses results - overall patient population (Figure B-19, of the MS).....100

Figure 5.3 Cost-effectiveness acceptability curves (Figure B-21, of the MS).....102

Figure 5.4 ERG deterministic sensitivity analyses of the manufacturer's base case results…….....103

Figure 6.1 The ICER as a function of time horizon.....110

Figure 6.2 Proportion of patients alive over time – ERG scenario assuming no continued treatment effect after 36 months.....112

Figure 6.3 Proportion of patients alive over time – ERG scenarios assuming that the OS curves converge over 12 or 18 month period. 113

List of abbreviations

5-FU	5-fluorouracil
AE	Adverse event
AFFIRM	Study of Aflibercept and Modified FOLFOX6 as First-Line Treatment in Patients with Metastatic Colorectal Cancer
AIC	Akaike information criterion
BIC	Bayesian information criterion
BNF	British National Formulary
BSA	Body surface area
BSC	Best supportive care
CEA	Cost-effectiveness analysis
CEAC	Cost-effectiveness acceptability curves
CHMP	Committee for Medicinal Products for Human Use
CI	Confidence interval
CRC	Colorectal cancer
CSR	Clinical study report
DH	Department of Health
DSA	Deterministic sensitivity analyses
DSU	Decision Support Unit
ECOG	Eastern Cooperative Oncology Group
EGFR	Epidermal growth factor receptor
EMA	European Medicines Agency
EORTC-QLQ	European Organisation for Treatment of Cancer Quality of Life Questionnaire
EQ-5D	EuroQol 5-dimension health questionnaire
ERG	Evidence Review Group
FACT	Functional Assessment of Cancer Therapy
FDA	Food and Drug Administration
FOLFIRI	Fluorouracil/leucovorin/irinotecan
FOLFOX	Folinic acid/5-fluorouracil/oxaliplatin
GI	Gastrointestinal
HR	Hazard ratio

HRG	Healthcare Resource Groups
HRQoL	Health related quality of life
HTA	Health Technology Assessment
ICER	Incremental cost-effectiveness ratio
IRC	Independent review committee
IV	Intravenous
IVRS	Interactive voice response system
KM	Kaplan-Meier
KOL	Key Opinion Leader
KRAS	Kirsten rat sarcoma viral oncogene homolog
mCRC	Metastatic colorectal cancer
MS	Manufacturer's submission
NHS	National Health Service
NICE	National Institute for Health and Clinical Excellence
NMB	Net monetary benefit
NR	Not reported
OS	Overall survival
PAS	Patient Access Scheme
PbR	Payment by Results
PD	Progressive disease
PFS	Progression free survival
PPS	Post-progression survival
PS	Performance status
PSA	Probabilistic Sensitivity Analysis
PSS	Personal social services
QALY	Quality-adjusted life year
RCT	Randomised controlled trial
RECIST	Response Evaluation Criteria in Solid Tumors
SAE	Serious adverse event
SD	Standard deviation
SE	Standard error
SPC	Summary of product characteristics
TTO	Time trade-off

UK	United Kingdom
VEGF	Vascular endothelial growth factor
VEGFR	Vascular endothelial growth factor receptor
VELOUR	Aflibercept Versus Placebo in Combination With Irinotecan and 5-FU in the Colorectal Cancer After Failure of an Oxaliplatin Based Regimen

1. SUMMARY

1.1 Critique of the decision problem in the manufacturer's submission

The manufacturer's submission (MS) evaluates aflibercept within its licensed indication "in combination with irinotecan/5-fluorouracil/folinic acid (FOLFIRI) chemotherapy in the treatment of adults with metastatic colorectal cancer (mCRC) that is resistant to or has progressed after an oxaliplatin-containing regimen". The population, intervention, comparator and outcomes in the MS are consistent with the National Institute for Health and Clinical Excellence (NICE) scope.

The description of the underlying health problem and overview of current service provision were appropriate and relevant to the decision problem under consideration. The NICE guideline on the diagnosis and management of colorectal cancer (CG131) recommends FOLFOX (folinic acid plus fluorouracil plus oxaliplatin) or XELOX (capecitabine plus oxaliplatin) as first-line chemotherapy, and then FOLFIRI as second-line chemotherapy for patients with advanced and metastatic colorectal cancer; therefore, aflibercept plus FOLFIRI as second-line chemotherapy is appropriate. However, in practice many patients are treated with irinotecan first-line and oxaliplatin second-line, as this sequence is associated with less cumulative toxicity (oxaliplatin is associated with moderate to severe neurological toxicity). Also in current clinical practice bevacizumab is often given in addition to these chemotherapy regimens, despite being not recommended by NICE.

The MS describes the mechanism of action of aflibercept, which demonstrates that there is a clear biological rationale for therapies such as aflibercept that target vascular endothelial growth factor (VEGF) pathways. The only other approved agent for the treatment of colorectal cancer that targets the VEGF pathway is bevacizumab.

1.2 Summary of clinical effectiveness evidence submitted by the manufacturer

The manufacturer conducted a systematic review evaluating the efficacy and safety of second-line treatments for mCRC. The evidence presented in the MS was derived from one good quality double-blind randomised controlled trial (RCT); the VELOUR trial, which compared aflibercept plus FOLFIRI with placebo plus FOLFIRI.

The trial found a small, but statistically significant, increase in overall survival (OS) for patients in the aflibercept group compared with patients in the placebo group; the difference in median OS was 1.44 months (13.5 months in the aflibercept group and 12.06 months in the placebo group). The survival increase was sustained, with a 22.3% survival probability in the aflibercept group compared with 12% in the placebo group at 30 months. However, there were very small numbers of patients included in the analysis at the later time points, which reduces the reliability of the longer-term results.

There was also a statistically significant increase in progression free survival (PFS) for the aflibercept group; the difference in median PFS between treatment groups was 2.23 months (6.9 months in the aflibercept group and 4.67 months in the placebo group). Response rate was also significantly higher in the aflibercept group compared with the placebo group (19.8% compared with 11.1%). Health related quality of life (HRQoL) was not assessed in the VELOUR trial.

Grade 3-4 adverse events were more frequent in the aflibercept group than the placebo group (83.5% compared with 62.5%).

[redacted]. Treatment-emergent adverse events led to permanent discontinuation of treatment in 26.8% patients in the aflibercept group and 12.1% patients in the placebo group.

The MS included an estimation of mean OS benefit based on an extrapolation of the data; 4.7 months. This was considerably longer than the median OS benefit reported (1.44 months), and also the restricted mean OS data supplied by the manufacturer upon request[redacted]. The Evidence Review Group (ERG) considers that this estimate may not be reliable.

1.3 Summary of the ERG's critique of clinical effectiveness evidence submitted

The MS presented a well conducted systematic review; the search strategy was adequate, inclusion screening and data extraction were undertaken in duplicate to reduce error and bias, adequate details of the included trial were presented and quality assessment was appropriate.

Superseded – see erratum

The trial included in the review was appropriate to the decision problem and was good quality. 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: in the UK between 2007 and 2009, an average 72% of colorectal cancer cases were diagnosed in people aged 65 years and over, whilst in the trial only 33.5% of the aflibercept group and 38.9% of the placebo group were aged over 65 years; the proportion of patients in the trial with an Eastern Cooperative Oncology Group (ECOG) performance status (PS) of 2 (2.2%) was lower than seen in other second-line clinical trials and standard UK practice; 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. 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.

The ERG considers that the estimates of mean OS may not be reliable due to the very low numbers of patients at risk at the later time points, particularly beyond 30 months. The continued separation of the survival curves is uncertain and the estimate of mean OS benefit (from extrapolation of the data) varies considerably, depending on which distribution is used (from 3.0 to [redacted] months). Truncating the extrapolation at 5 years (which seems a more reasonable assumption to the ERG, since it is unlikely that the treatment benefit would extend beyond 5 years), rather than 15 years, improved consistency across the different models, and reduced the OS benefit [redacted] (data supplied by manufacturer upon request). Finally, the ERG does not entirely agree that the log-cumulative hazard plots for OS support the manufacturer's argument against using a proportional hazards approach to generating the survival curves, and further would suggest that it would be reasonable to assume that the survival curves would converge before 5 years.

1.4 Summary of cost effectiveness evidence submitted by the manufacturer

The manufacturer undertook a literature review of economic evaluations in mCRC. No previously published cost-effectiveness analyses of aflibercept were reported by the manufacturer. A *de novo* model was provided by the manufacturer. The manufacturer's model was a three state Markov model with a time horizon of 15 years. The transition probabilities for progression, overall survival and discontinuation of treatment were based on the VELOUR trial. Health-related quality-of-life

Superseded – see erratum

(HRQoL) was estimated from a separate, cross-sectional study of mCRC patients. Resource use was obtained from a retrospective observational study in mCRC patients and a survey of clinical oncologists from the UK, the Key Opinion Leader (KOL) survey. National unit costs from published sources were applied to resource use. Adverse event rates were calculated from the VELOUR trial. The manufacturer reports mean costs and quality-adjusted life-years (QALYs) and the incremental cost-effectiveness ratio (ICER) for the overall population and two subgroups, a pre-specified liver metastases only' subgroup and a *post-hoc* subgroup that excluded patients who had received adjuvant chemotherapy only. The ICER for the overall population was reported to be £36,294 per QALY. Results were reported to be sensitive to the selection of parametric OS distributions and the utility of the progressive disease health state. The probability of cost-effectiveness was reported to be 89.5% at £50,000 per QALY and 22% at £30,000 per QALY. All analyses presented by the manufacturer (with the exception of budget impact estimation) were based on the acquisition cost for aflibercept based on a patient access scheme (PAS). 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.

1.5 Summary of the ERG's critique of cost effectiveness evidence submitted

The ERG concluded that the manufacturer's literature review was well conducted. One abstract has been published on the cost-effectiveness of aflibercept versus bevacizumab in the US since the manufacturer's literature search was completed. However, the ERG did not consider that this new study provided information relevant to the decision problem. The structure of the *de novo* model provided by the manufacturer was similar to previous models in mCRC. The 15 year time horizon of the model meant that the assumptions used to extrapolate results from the VELOUR trial (median follow-up time of 22 months) over a longer time horizon was central to the cost-effectiveness results. The major data sources of the model were studies in populations that appear to be younger and arguably healthier than the second-line mCRC patient population treated in the UK clinical practice. The economic evaluation was for patients using second-line treatment after an oxaliplatin-containing regimen, the ERG noted that this patient group is narrower than the licensed indication, which does not specify second-line use. The use of FOLFIRI alone as the only comparator in the evaluation was considered to be appropriate. The perspective, time horizon and discounting in the model were appropriate. The economic modelling of progression and OS reflected the parametric survival modelling reported in the clinical effectiveness sections. In the base case analysis the manufacturer

Superseded – see erratum

applied independent parametric survival functions to the separate arms of the VELOUR trial and continued to apply these functions during the extrapolation period. 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. The manufacturer's model appears to have been developed to also allow an assumption of proportional hazards to be made (i.e. assuming a constant relative effect over the entire model horizon). However, the results of this scenario were not presented within the manufacturer's report. 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. The extrapolation of the OS curves was the key cost-effectiveness driver and a major source of uncertainty. The ERG did not consider that the manufacturer had sufficiently explored the uncertainty around the extrapolation of the OS curves with the scenarios they presented. The ERG consider that alternatives should be explored using different assumptions, such as, assuming the treatment effects continue in a similar manner to the observed period, assuming that any benefits immediately cease at the end of the observed period and assuming that any benefits decrease over the extrapolation period. Only the first of these is explored by the manufacturer.

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, secondly, the cycle length adjustment was undertaken twice for AE decrement, thirdly, no adjustment of utility estimates were made to account for the impact of patient ageing in the model, and finally, the patients in the model represented a younger population than those in clinical practice in the UK and no adjustment had been made for this. The ERG also identified several model parameters that did not seem to be well supported by the evidence. The manufacturer's utility value for progressive disease appeared high compared to values used in previous analyses in mCRC or compared to utilities collected in the general UK population for patients 65-74 years old. The ERG also considers that the administration cost of aflibercept has been underestimated. The manufacturer assumed there would be no additional administration costs from adding aflibercept to FOLFIRI. However, previous HTA reports used to support NICE appraisals (HTA [1], have included an additional pharmacy preparation cost (£15) for each additional infusional treatment used in combination chemotherapy. In the model the manufacturer also applied the median resource use from the survey of UK oncologists (KOL survey). However, mean values are most commonly used in cost-effectiveness analysis, in order to estimate expected costs and benefits.

Superseded – see erratum

1.6 ERG commentary on the robustness of evidence submitted by the manufacturer

1.6.1 Strengths

The description of the underlying health problem and overview of current service provision were appropriate and relevant to the decision problem under consideration. The evidence presented for the clinical effectiveness of aflibercept was identified through a well conducted systematic review, which used established methodology to reduce the risk of error and bias. The evidence for the effectiveness and safety of aflibercept comprised one good quality RCT.

The economic evaluation was based on a *de novo* model which the ERG considers to be the most relevant evidence to inform the decision model. . Many of the model parameters were based on a good quality RCT. A well conducted systematic review provided additional information on model parameters. Uncertainty was explored by the manufacturer through extensive sensitivity analyses.

1.6.2 Weaknesses and areas of uncertainty

The demographic and disease characteristics of the single trial's participants suggest that they were potentially younger and healthier than the population eligible for aflibercept plus FOLFIRI in UK practice. Therefore, the level of benefit achieved in the trial may not be realised in clinical practice. 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. The ERG considers that the estimates of mean OS are very uncertain. There were no HRQoL data presented in the clinical section of the MS, which the manufacturer acknowledges as a limitation of the clinical evidence base of the intervention.

As discussed in Section 1.5 the ERG identified some problems with the model programming that resulted in incorrect calculations of acquisition and administration costs and the utility decrements for AE. The manufacturer also excluded age dependent utilities from the model. The ERG found that some of the model parameters were not well supported by the evidence: the utility for the progressive disease health state, the administration costs and median resource use values. Importantly the uncertainty around the OS estimates was not adequately explored. The ERG was provided with no information on the manufacturer's PAS, so the implementation of the PAS in the model could not be verified.

Superseded – see erratum

1.7 Summary of exploratory and sensitivity analyses undertaken by the ERG

The ERG defined an alternative 'base case' having corrected the calculations for acquisition and administration costs and utility decrements of adverse events. The ERG alternative base case also included an age adjusted HRQoL, a lower estimate of the utility for progressive disease (0.60), a small additional cost to the administration of aflibercept (£15) and used the mean resource use from the KOLUK oncologist survey. These changes increased the ICER to £54,368 per QALY.
The ERG tested a number of alternative OS assumptions. In each of the ERG analyses the manufacturer's base case treatment effect was assumed to hold until either 30 or 36 months follow-up. These time points were identified as being highly uncertain in the manufacturer's OS piecewise model due to the small numbers of patients still at risk. The ERG explored two assumptions other than the manufacturer's base case which assumed that the treatment effect of aflibercept increased over time. Firstly the ERG tested an assumption where the treatment effect stopped at 30 or 36 months after which the risk of death was the same between arms. Secondly the ERG tested an assumption where the treatment effect stopped at 30 or 36 months after which the risk of death in the aflibercept + FOLFIRI arm was higher than the FOLFIRI alone arm. The effect of this assumption was to allow for the survival curves to converge over the period of extrapolation. At the point the survival curves crossed all patients were subsequently assumed to have the same mortality regardless of initial treatment. The second assumption was undertaken to correspond with the increased progression found in the aflibercept + FOLFIRI arm resulting in the PFS curves converging at about 12 months. The ERG's exploration of the OS assumptions resulted in ICERs ranging from £62,894 to £92,089 per QALY.

The ERG undertook further analyses in the subgroups presented by the manufacturer. Combining the ERG's OS assumptions and the ERG's alternative base case the ICER of aflibercept ranged from £46,576 to £58,257 per QALY in the liver metastases only subgroup and from £57,224 to £80,187 per QALY in the excluding adjuvant chemotherapy subgroup.

1.8 Conclusions or key issues

Evidence from one good quality RCT demonstrates a statistically significant, but clinically small (difference in median OS 1.44 months), increase in OS and PFS with aflibercept plus FOLFIRI,

compared with FOLFIRI alone. However, grade 3-4 adverse events are common (83.5% in the aflibercept group compared with 62.5% in the placebo group), leading to discontinuation of treatment in a large number of patients (26.8% in the aflibercept group compared with 12.1% in the placebo group). There are currently no data on HRQoL for patients taking aflibercept.

No previously published cost-effectives studies were pertinent to the decision problem. The manufacturer's *de novo* cost-effectiveness analysis met the criteria for the NICE reference case. The manufacturer reported a base case ICER of £36,294 per QALY for aflibercept plus FOLFIRI compared with FOLFIRI alone. After correcting the model programming and updating the model to include the ERG's preferred parameter estimates the ICER from the ERG's alternative base case was £54,368 per QALY. In addition, the ERG also considers that the uncertainty surrounding the extrapolation period is not fully characterised by the scenarios explored by the manufacturer. Additional scenarios related to the extrapolation of OS undertaken by the ERG resulted in ICERs between £62,894 to £92,089 per QALY. Although these scenarios applied more conservative assumptions than those employed by the manufacturer, the ERG considers that these are equally plausible to those presented by the manufacturer. Furthermore, the wide range of ICER estimates resulting from the different assumptions clearly highlights that this is the key driver of cost-effectiveness results and also that the high uncertainty surrounding the extrapolation period stems largely from the high censoring evident in the current data beyond 30 months. Hence, the ERG considers the magnitude of extrapolation of the mean OS benefit applied in the economic model (4.7 months) to be both highly uncertain and also overly optimistic compared to the observed median OS benefit reported (1.44 months).

Superseded – see erratum

2 BACKGROUND

2.1 Critique of manufacturer's description of underlying health problem

The description of the underlying health problem in the manufacturer's submission (MS) is adequate and relevant to the decision problem under consideration.

The MS states that the 5-year survival rate for patients receiving second-line metastatic colorectal cancer (mCRC) therapy is estimated at 10%, and observed median OS ranges between 10.5 and 18 months; the MS references Seymour *et al*., 2011,[2] Jonker *et al*., 2007,[3] Peeters *et al*., 2010[4] and Kozloff *et al*., 2011.[5] The study by Seymour *et al*.,[2] was conducted in the UK, so is likely to be the most reflective of the UK population; the median OS for this trial was 10.5 months.

2.2 Critique of manufacturer's overview of current service provision

The manufacturer's overview of current service provision is appropriate and relevant to the decision problem under consideration. The appropriate NICE guideline on the diagnosis and management of colorectal cancer (CG131) is referenced.[6] This guideline recommends FOLFOX (folinic acid plus fluorouacil plus oxaliplatin) or XELOX (capecitabine plus oxaliplatin) as first-line chemotherapy, and then FOLFIRI as second-line chemotherapy for patients with advanced and metastatic colorectal cancer; therefore, aflibercept plus FOLFIRI as second-line chemotherapy is appropriate. However, the ERG's clinical advisor informed the ERG that in practice many patients are treated with irinotecan first-line and oxaliplatin second-line, as this sequence is associated with less cumulative toxicity, since first-line chemotherapy is given for longer than second-line chemotherapy; a study comparing FOLFIRI followed by FOLFOX6 with the reverse sequence found similar efficacy for both sequences but different toxicity profiles, grade 3-4 toxicities (particularly neurotoxicity) were more frequent when FOLFOX6 was given first-line.[7] The MS correctly states that, to date, no biologic therapies have demonstrated cost-effectiveness and been approved by NICE in the second-line setting (TA242).[8]

In standard UK practice, patients are often given intermittent first-line palliative chemotherapy,[9] and patients who progress after a treatment break are likely to be offered repeat treatment with the first-line chemotherapy regimen; the criteria for switching to second-line therapy is progression on treatment or symptomatic progression within six to eight weeks of completion of first-line chemotherapy. Therefore, patients may undergo several cycles of first-line palliative therapy before

progressing to second-line therapy. Intermittent first-line chemotherapy and criteria for switching to second-line therapy are not described in the MS.

The MS describes the mechanism of action of aflibercept, which is briefly summarised below. Angiogenesis (formation of new blood vessels) plays a critical role in the growth and migration of solid tumours. The most important signalling factors in angiogenesis are the vascular endothelial growth factor (VEGF) ligands, which include VEGF-A, VEGF-B, VEGF-C, VEGF-D, and placental growth factor (PlGF), which exert their effect through receptors, including VEGFR-1, VEGFR-2, and VEGFR-3. Therefore, there is a clear biological rationale for therapies that target VEGF pathways.

Aflibercept is a fully human, recombinant fusion protein that consists of VEGFR extracellular domains (VEGFR-1 and VEGFR-2) fused to the Fc portion of human immunoglobulin G1. Aflibercept acts as a decoy receptor that blocks the VEGF pathway by preferentially binding to VEGF ligands in a novel “trap” mechanism. Once bound to aflibercept, VEGF ligands are unable to bind to and activate their cognate receptors.

The only other approved agent for the treatment of colorectal cancer that targets the VEGF pathway is bevacizumab, which is a monoclonal antibody against VEGF-A. The manufacturer states that aflibercept binds to VEGF-A with approximately 100-fold higher affinity than bevacizumab. In addition to binding to VEGF-A, aflibercept also binds to VEGF-B and PlGF.

3 CRITIQUE OF MANUFACTURER'S DEFINITION OF DECISION PROBLEM

3.1 Population

The population in the MS matches that specified in the NICE scope, "people with metastatic colorectal cancer (mCRC) that is resistant to or has progressed following prior oxaliplatin-based chemotherapy". This is within the licensed indication of aflibercept, which is, "for the treatment of adults with mCRC that is resistant to or has progressed after an oxaliplatin-containing regimen".

Whilst the NICE scope did not identify any subgroups to be considered, the MS considers the following two subgroups: patients with metastases confined to the liver, and a *post-hoc* subgroup excluding patients progressing on or within six months of oxaliplatin-based adjuvant treatment, i.e. those who had received adjuvant therapy only. The MS states that the subgroup of patients with liver metastases only has been identified in previous NICE guidance as a recognised clinical subgroup in mCRC. The MS also states that rapid relapse after adjuvant therapy is typically associated with very aggressive disease and poor prognosis, therefore the subgroup of patients who progressed on or within six months of oxaliplatin-based adjuvant treatment may have tumours with genetic changes that make them less sensitive to chemotherapy.

3.2 Intervention

Aflibercept, in combination with irinotecan, 5-fluorouracil (5-FU) and folinic acid (FOLFIRI) chemotherapy, has a UK marketing authorisation for the treatment of adults with mCRC that is resistant to or has progressed after an oxaliplatin-containing regimen. The intervention described in the MS matches the licence and that specified in the NICE scope: aflibercept + FOLFIRI.

Aflibercept is administered as an intravenous (IV) infusion over one hour every two weeks (cycle) at a dose of 4 mg/kg of body weight alongside a standard FOLFIRI regimen. The treatment cycle is repeated every two weeks until disease progression or unacceptable toxicity. Following administration of aflibercept, all patients will receive the FOLFIRI regimen, administered as follows: irinotecan 180 mg/m^2 IV infusion over 90 minutes and leucovorin (dl racemic) 400 mg/m^2 IV infusion over two hours at the same time on day one using a Y line, followed by 5-FU 400 mg/m^2 IV bolus, followed by 5-FU 2,400 mg/m^2 continuous IV infusion over 46 hours. The MS states that in practice

FOLFIRI doses may be adapted and this dosing schedule is only indicative. This FOLFIRI regimen is the same as the standard FOLFIRI regimen used in the UK.

In addition to aflibercept and FOLFIRI, patients may also be prescribed therapies to manage adverse reactions associated with aflibercept and FOLFIRI, such as intravenous or subcutaneous atropine (0.25-1 mg) for cholinergic adverse effects, prophylactic anti-emetic therapy, granulocyte-colony stimulating factor for grade 3 or 4 neutropaenia and antihypertensive therapies.

3.3 Comparators

The comparator specified in the NICE scope was stated as irinotecan + fluorouracil-based therapy. The comparator specified in the decision problem was FOLFIRI (irinotecan + fluorouracil and leucovorin). The rationale for the difference in the stated comparator was that FOLFIRI is a regimen that contains irinotecan + fluorouracil-based therapy. Therefore, the comparator described in the MS matches that described in the NICE scope. The NICE guideline on the diagnosis and management of colorectal cancer (CG131) recommends FOLFIRI as second-line chemotherapy for patients with advanced and metastatic CRC;[6] therefore, FOLFIRI is an appropriate comparator. NICE does not recommend cetuximab (on its own or in combination with chemotherapy) or bevacizumab in combination with a fluoropyrimidine-based chemotherapy or panitumumab on its own as second-line chemotherapy for mCRC as cetuximab and panitumumab were not considered a cost-effective use of NHS resources and there was insufficient evidence of the clinical efficacy of bevacizumab.[10]

Although cetuximab, panitumumab and bevacizumab are not considered as direct comparators of aflibercept for this NICE Single Technology Appraisal, the ERG has included some brief information on these agents in order to provide some context for the clinical efficacy of aflibercept.

The epidermal growth factor receptor (EGFR) inhibiting monoclonal antibody agents cetuximab and panitumumab are indicated for the second-line treatment of mCRC, but only for those patients with wild-type KRAS tumour status. KRAS status is not of relevance to the efficacy of aflibercept and therefore these drugs cannot be considered direct comparators of aflibercept. Also, whilst cetuximab is licensed for use in combination with irinotecan-based chemotherapy, it is generally reserved for use as a single agent in patients who have failed oxaliplatin- and irinotecan-based therapy and who are intolerant to irinotecan. A non-systematic search by the ERG failed to identify any randomised controlled trials (RCTs) that compared cetuximab + FOLFIRI with FOLFIRI alone in second-line treatment.

Regarding panitumumab, of interest to the appraisal of aflibercept is a trial of panitumumab that compared panitumumab + FOLFIRI with FOLFIRI alone in second-line mCRC.[4] In a sub-group of patients with wild-type KRAS (but not the whole trial population) there was a statistically significant treatment benefit in terms of PFS (median difference 2.0 months, HR 0.73 (95% CI 0.59 to 0.90)) but not OS (median difference 2.0 months, HR 0.85 (95% CI 0.70, 1.04)); the lack of statistical significance may be due to the small sample size.

Bevacizumab is, like aflibercept, a VEGF inhibitor, the clinical benefit of which is independent of KRAS status.[11, 12] Under its European Medicines Agency (EMA)/UK licence, "bevacizumab in combination with fluoropyrimidine-based chemotherapy is indicated for treatment of adult patients with metastatic carcinoma of the colon or rectum." The licence places no restrictions on the use of bevacizumab in mCRC regarding previous therapy. Therefore, whilst bevacizumab does not have a specific label indication for use in combination with FOLFIRI, its use in combination with FOLFIRI is covered by the licence, and thus bevacizumab + FOLFIRI as second-line therapy is a direct comparator for aflibercept + FOLFIRI in its licensed indication. The product licence for bevacizumab was supported by evidence from four trials as first-line therapy and two trials as second-line therapy.[13-15]

No RCT of bevacizumab + FOLFIRI compared with FOLFIRI alone as second-line therapy has been identified. A non-systematic search by the ERG identified a number of publications that report experience with this specific combination in second-line mCRC. Of particular relevance was a meta-analysis based on all studies identified from a search of MEDLINE and EMBASE up to October 2012.[16] This pooled analysis assessed bevacizumab + FOLFIRI as second-line therapy in mCRC patients pre-treated with oxaliplatin. None of the included studies contained a randomised comparison with FOLFIRI alone and most were small single arm retrospective studies.

As stated in the MS, aflibercept "is the only anti-angiogenic agent to show an OS benefit in the second-line setting in combination with FOLFIRI".

The ERG searched for trials of other potentially relevant comparators (RCTs of other therapies used in combination with FOLFIRI in the second-line treatment of mCRC). The following RCTs were found:

One RCT of trebananib (AMG 386) in combination with FOLFIRI compared with placebo and FOLFIRI.[17]
One terminated RCT of zibotentan in combination with FOLFIRI compared with placebo and FOLFIRI.[18]

However, none of the trials were considered to be relevant to compare against aflibercept in a formal indirect comparison; the trial of trebananib found no benefit of trebananib over FOLFIRI alone, and the trial of zibotentan was terminated early due to lack of efficacy (the trial concluded that zibotentan does not have anti-cancer activity in combination with FOLFIRI in patients with colorectal cancer treated with previous FOLFOX).

Whilst a formal indirect comparison was not considered appropriate, this report will discuss the clinical effectiveness and adverse effects of bevacizumab (in Section 4.4.1), as this is the only other licenced biological therapy that works in the same way as aflibercept, by targeting the VEGF pathway for the treatment of colorectal cancer. The ERG's clinical advisor informed the ERG that in clinical practice bevacizumab is often used in combination with FOLFOX as first-line chemotherapy for patients with mCRC, and is also used alongside FOLFIRI in second-line chemotherapy.

3.4 Outcomes

All of the outcomes specified in the NICE scope were considered in the MS; overall survival (OS), progression-free survival (PFS), response rate, adverse effects of treatment, and health-related quality of life (HRQoL). However, there were no HRQoL data presented in the clinical section of the MS, which the manufacturer acknowledges as a limitation of the clinical evidence base of the intervention.

PFS and OS in mCRC

In trials evaluating the efficacy of a treatment for cancer, OS and PFS are common endpoints. OS is an accurate measure of the time of death, whereas PFS is a more complex measure of disease progression or death.[19, 20] PFS is increasingly being used as the primary outcome in trials of cancer therapy, as a surrogate for OS, mainly due to the long follow-up time needed to collect mature OS data, but also because of the increasing duration of post-progression survival (PPS) due to the availability of second- and third-line therapies. There are several disadvantages to the use of PFS, primarily (but not solely) due to the complex relationship between PFS and OS; the ability of PFS to

accurately predict OS across populations and/or treatments may vary considerably.[21, 22] A benefit in terms of PFS will not necessarily translate into an increase in OS; it may be associated with either no benefit or a reduction in OS. Examples as to when this could occur are where there is long term-toxicity associated with the intervention, where different treatments have different resistance profiles, or where biological changes lead to increased metastatic potential.[21] There are, however, some advantages to the use of PFS. In addition to the shorter timelines for data collection, PFS is not affected by cross-over or subsequent therapies and it is generally based on objective and quantitative assessment.[23] Furthermore, patients value PFS so it is a valued outcome in its own right.[24]

The most recent systematic review of the association between PFS and OS was conducted by the Decision Support Unit (DSU), which investigated this relationship for all types of cancer.[19] 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.

The DSU report included five studies in mCRC, one of which investigated second-line therapy.[25] The analysis of the full data set (62 studies, 23,527 patients) yields similar results to other studies, with a high positive correlation between median PFS and median OS (r=0.87, R^2=0.48). However R^2 was higher for first-line (0.54) than for second-line therapy (0.38) and the slope of the regression line was less in the second-line model indicating a larger PFS effect is needed in second-line therapy to achieve the same gain in OS as in first-line.

Since the searches for the DSU report were completed a number of further studies of the relationship between PFS, PPS and OS in mCRC have been published.[24, 26, 27] Whilst these studies also found a strong correlation between PFS and OS and confirmed PFS as an acceptable surrogate endpoint for cytotoxic chemotherapies, an analysis of bevacizumab plus chemotherapy was based on heterogeneous data and yielded weak correlation between the endpoints themselves and wide confidence intervals.[26]

Examination of not only the association between OS and PFS but also the association between PPS and OS in first-line therapy[27] found that whilst median PFS was moderately correlated with median OS (r=0.64, R^2=0.43), PPS was strongly correlated with median OS (r=0.88, R^2=0.80). Thus PPS

explained much more of the variation in OS than did PFS, indicating that post-progression therapy has a greater effect on OS than the therapy under investigation. The authors conclude that "this strong association of PPS with OS is the result of the increasing number of active compounds available also for second- or third-line chemotherapy in advanced CRC".

Whilst there are theoretical reasons to suppose that gains in OS may exceed gains in PFS, the available clinical evidence suggests that for second-line therapies in mCRC, OS is highly correlated with PFS, but may be more so with PPS and that this is most likely influenced by post-progression therapies rather than the therapy under study.

3.5 Other relevant factors

A PAS was approved by the Department of Health and is applied to the cost-effectiveness analyses; the ERG requested further information about the PAS, but did not receive any details.

4 CLINICAL EFFECTIVENESS

This section contains a critique of the methods of the systematic review presented in the MS, followed by a description and critique of the trials included in the review, including a summary of their quality and results. The ERG's conclusions on the clinical effectiveness of aflibercept, in combination with irinotecan and fluorouracil-based therapy, for the treatment of mCRC which has progressed following prior oxaliplatin-based chemotherapy, are presented at the end of this section.

4.1 Critique of the methods of review(s)

The MS described a systematic review of studies evaluating second-line treatment of mCRC. One RCT (the VELOUR trial) compared aflibercept in combination with FOLFIRI with placebo and FOLFIRI in patients with mCRC previously treated with an oxaliplatin-based regimen.

The MS also presented the results of a meta-analysis of aflibercept adverse events; however, minimal details of the meta-analysis were presented. The meta-analysis included three RCTs; the VELOUR trial, a trial of patients with non-small cell lung cancer (the VITAL trial) and a trial of patients with metastatic pancreatic cancer (the VANILLA trial). The ERG requested the full report of the meta-analysis of aflibercept adverse events, but this was not provided by the manufacturer.

4.1.1 Search strategy

The MS described the search strategies used to identify relevant clinical effectiveness studies about the use of aflibercept for the treatment of mCRC previously treated with an oxaliplatin containing regimen. Search strategies were only briefly described in the main body of the submission, however full details were provided in the Appendices.

The electronic databases MEDLINE and MEDLINE In-Process (via PubMed), EMBASE and Cochrane Central Register of Controlled Trials (CENTRAL) were searched to identify clinical studies on the use of aflibercept in the second-line treatment of mCRC. In addition to this, abstracts of conference proceedings from the American Society of Clinical Oncology (ASCO) and the European Society of Medical Oncology (ESMO) were reviewed.

Searches were conducted on 19 December 2012. Search strategies for each database were documented in Tables 10-1 to 10-4. The searches covered the period 1 January 1992 to December 2012, were

Superseded – see erratum

limited to English language publications, and excluded animal studies. In addition, the EMBASE search strategy excluded all ‘conference’ publication types and the ‘review’ publication type. The manufacturer did not explain why the search was limited by date range and to only English language publications.

Overall the searches were appropriate and well documented, and included the use of both subject indexing terms (MeSH and EMTREE) and free text searching. Field searching, Boolean operators and truncation were used where required. All the databases required by NICE were searched, though only CENTRAL was searched in the Cochrane Library when it might have been useful to have searched the Cochrane Database of Systematic Reviews (CDSR), the Database of Abstracts of Reviews of Effects (DARE) and the Health Technology Assessment (HTA) database. Information about the EMBASE service provider was not provided, but was subsequently provided via the Points of Clarification letter.

There were some issues with the use of MeSH and EMTREE terms. The PubMed search strategy included both MeSH and Major MeSH index terms; Major MeSH terms are redundant when used alongside the equivalent MeSH terms. Further, the MeSH terms used (‘Colorectal Neoplasms’, ‘Colonic Neoplasms’ and ‘Rectal Neoplasms’) were not exploded, and so additional MeSH terms found further down the MeSH hierarchy would not have been searched for. In the EMBASE search strategy MESH terms are used instead of EMTREE terms in search line #1.

Other issues included: a missing Boolean operator (OR) in search line #2; a redundant search line (line #11); and unnecessary repetition of search terms (lines #4 and #36). The drug name aflibercept (and related terms) was not included in the search strategies and so it is possible that potentially useful records were not retrieved. The search strategies used in the manufacturer's submission were limited to randomised controlled trials and phase II and phase III trials, however a search for other study designs such as cohort or case control studies may have provided useful supplementary information about safety. It is not clear if the methodological search filters used in PubMed, EMBASE and the Cochrane Library were derived from validated search filters. The addition of the following EMTREE terms would have improved the filter used in EMBASE: ‘Randomized Controlled Trial’ and ‘Controlled Clinical Trial’.

Despite the issues identified above, the searches were appropriate and comprehensive, and included the use of both subject indexing terms and free text searching. Field searching, Boolean operators and truncation were used where required. It is unlikely that any relevant studies have been missed.
No separate search strategy was undertaken to identify relevant studies for the meta-analysis of the adverse effects of aflibercept.

4.1.2 Inclusion criteria

RCTs of all current second-line chemotherapy regimens (including, but not limited to, bevacizumab, irinotecan, FOLFIRI, FOLFOX4, FOLFOX6, capeOX and XELOX) compared with placebo, best supportive care or the same or a different second-line chemotherapy regimen, for adult patients with mCRC were eligible for inclusion in the review. Trials had to assess outcomes relating to survival, progression, response, adverse effects or quality of life to be included in the review. Only English language articles were eligible for inclusion.

The inclusion criteria were appropriate for a systematic review of all second-line chemotherapy regimens for adult patients with mCRC. It appears that at the study selection stage the inclusion criteria were narrowed down so that the only intervention eligible for inclusion in the review was aflibercept.

Two reviewers independently assessed the identified titles and abstracts and then full text articles against the inclusion and exclusion criteria, reducing the potential for error or bias. Whilst only English language articles were selected for the review, creating the potential for language bias, it is unlikely that any relevant studies of aflibercept were excluded.

A flow chart of the study selection process was presented in the MS; however, no details were presented of the reasons for exclusion of the citations or full-text publications that did not meet the inclusion criteria. The flow chart stated that 43 publications of 30 RCTs were included in the review. However, only one RCT evaluated the intervention under appraisal (aflibercept), and the review included only one RCT (the VELOUR trial).

No inclusion criteria were presented for the meta-analysis of the adverse effects of aflibercept.

4.1.3 Data extraction

Two reviewers independently extracted data from the included study, reducing the potential for error or bias. The MS presented adequate data from the VELOUR trial.

Very little data from the VITAL and VANILLA trials were presented. The ERG requested full adverse event results from the VITAL and VANILLA trials, but these were not provided by the manufacturer.

4.1.4 Quality assessment

The quality of the VELOUR trial was assessed using appropriate criteria specific to RCTs, and a table of the quality assessment results was presented as Table B7. Quality assessment results were checked by the ERG.

It is unclear whether there was any quality assessment of the VITAL and VANILLA trials.

4.1.5 Evidence synthesis

No meta-analysis of efficacy outcomes was undertaken as only one study of aflibercept in the second-line treatment of mCRC was identified in the systematic review.

A meta-analysis of aflibercept adverse events was presented, including the VELOUR trial, the VITAL trial and the VANILLA trial.

4.1.6 Conclusions from critique of systematic review methods

The search strategy was adequate; no relevant studies of aflibercept for the second-line treatment of mCRC were overlooked. However, the flow chart of the study selection process did not present the reasons for exclusion of the trials that did not meet the inclusion criteria. Inclusion screening and data extraction were undertaken in duplicate to reduce error and bias. Adequate details of the included trial were presented, including an appropriate quality assessment; the trial was good quality.

Minimal details of the meta-analysis of aflibercept adverse events were presented in the MS; no separate search strategy to identify relevant studies was undertaken. It is unclear whether there was any quality assessment of the VITAL and VANILLA trials. The ERG requested the full report of the

meta-analysis of aflibercept adverse events and full adverse event results from the VITAL and VANILLA trials, but these were not provided by the manufacturer.

4.1.7 Ongoing studies

The MS states that the following study involving aflibercept is ongoing and expected to report results in the next 12 months. This study will provide additional data on the safety of aflibercept and HRQoL of the included patients in order to capture utility values not collected during the VELOUR trial; however, the study will not compare HRQoL between patients taking aflibercept and those not taking aflibercept.

- Study AFLIBC06097 (ASQoP) is a prospective phase 3b/4 single arm trial assessing safety (treatment-emergent adverse events) and HRQoL using the EORTC QLQ-C30 (version 3), EORTC QLQ-CR29 (disease specific supplement to QLQ-C30) and EQ-5D (version 4).

In addition, the ERG identified the following ongoing study of aflibercept.

- 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.

4.2 Critique of trials of the technology of interest, their analysis and interpretation (and any standard meta-analyses of these)

4.2.1 Trials included in the review

One multicentre double-blind parallel-group RCT of aflibercept in combination with FOLFIRI compared with placebo and FOLFIRI was included in the review; the VELOUR trial. The trial randomised 1,226 patients between November 2007 and March 2010 on a 1:1 basis to intravenous (IV) aflibercept (4mg/kg) plus FOLFIRI or IV placebo plus FOLFIRI. Hereafter, the aflibercept plus FOLFIRI group will be referred to as the aflibercept group and the placebo plus FOLFIRI group will be referred to as the placebo group. The trial was conducted in 176 centres in 28 countries: Argentina, Australia, Austria, Belgium, Brazil, Chile, Czech Republic, Denmark, Estonia, France, Germany, Greece, Italy, Korea, Netherlands, New Zealand, Norway, Poland, Puerto Rico, Romania, Russian Federation, South Africa, Spain, Sweden, Turkey, Ukraine, United Kingdom (n=99) and United States. The trial was directly relevant to the decision problem.

The design and methods of the VELOUR trial are summarised in Table 4.1, baseline participant characteristics are presented in Table 4.2.

Table 4.1 Design and methods of the VELOUR trial

Study details	**VELOUR trial (study EFC10262)**
Design	Randomised, double-blind, placebo controlled
Duration of study	Median follow-up for intent-to-treat population: 22.28 months
Method of randomisation	Interactive Voice Response System (IVRS) based on a permuted-block randomisation, stratified according to prior therapy with bevacizumab (yes or no), and Eastern Cooperative Oncology Group (ECOG) performance status (PS) (0, 1, or 2)
Method of blinding (care provider, patient and outcome assessor)	Aflibercept and placebo were supplied in indistinguishable sealed vials in identical boxes corresponding to patient kits; during the course of the study, an external statistician (independent from the Sponsor) performed unblinded safety and efficacy interim analyses for the purpose of data review by the Data Monitoring Committee
Intervention(s)	Aflibercept 4 mg/kg + FOLFIRI (n = 612; 52 in UK)
Comparator(s)	Matched placebo + FOLFIRI (n=614; 47 in UK)
FOLFIRI regimen	Irinotecan 180 mg/m^2 IV infusion over 90 minutes and leucovorin 400 mg/m^2 IV infusion over 2 hours at the same time in bags using a Y-line, followed by 5-FU 400 mg/m^2 IV bolus given over 2-4 minutes, followed by 5-FU 2,400 mg/m^2 continuous IV infusion over 46 hours
Stopping rules	Patients were to be treated until occurrence of disease progression or unacceptable toxicity according to physician judgement. Patients could stop one drug but still continue treatment with the other components of the regimen. If FOLFIRI was permanently discontinued, then aflibercept or placebo could be continued until disease progression, unacceptable toxicity, or patient's refusal of further treatment. The end-of-study treatment date was the date of the last aflibercept or placebo administration. If aflibercept or placebo was permanently discontinued, then FOLFIRI could be continued until disease progression, unacceptable toxicity, or patient's refusal of further treatment. The end-of-study treatment date was the date of the last FOLFIRI administration. Any anti-cancer treatment administered after the last study treatment administration until death or until study termination (whichever came first) was documented
Primary outcome	Overall survival (OS) (defined as the time interval from randomisation to death from any cause)
Secondary outcomes	Progression free survival (PFS) (defined as the time interval from randomisation to the first observation of disease progression (according to independent review committee review) or death from any cause) Objective response (complete response and partial response (according to Response Evaluation Criteria In Solid Tumours (RECIST) criteria version 1.0)) Treatment-emergent adverse events and laboratory abnormalities
Duration of follow-up	Following documentation of progressive disease, patients were to be followed for survival status every 2 months until death or withdrawal of patient consent or until the cut-off date for final analysis of OS was reached, whichever came first

Study details	**VELOUR trial (study EFC10262)**
	Patients who discontinued treatment before disease progression were followed every 6 weeks for PFS. Once progression was documented (according to investigator evaluation), surviving patients were followed every 2 months for survival until death or study cut-off After data cut-off for primary analysis of OS, patients who were still receiving study treatment and patients who were alive and had completed study treatment, and were experiencing ongoing serious adverse events (SAEs) or new/ongoing related adverse events (AEs), were to be followed for safety purposes until stabilisation or resolution for a maximum of 9 months after the study cut-off
Patient inclusion criteria	Aged 18 years or older ECOG PS of 0 to 2 Histologically or cytologically proven colorectal adenocarcinoma Metastatic disease not amenable to potentially curative treatment Measurable or nonmeasurable disease (as per Response Evaluation Criteria In Solid Tumors [RECIST] criteria) Only one prior chemotherapeutic regimen for metastatic disease, which had to be an oxaliplatin-containing regimen. Although patients were to have documented progression while on or after completion of a single, prior, oxaliplatin-containing regimen, they were not selected for the timing of their progression. Patients who experienced relapse on or within 6 months of completion of oxaliplatin-based adjuvant therapy were eligible Prior bevacizumab was permitted, prior irinotecan was not Randomisation occurred 28 days or more after cessation of prior radiotherapy or chemotherapy and 30 days or more after ending participation in another clinical trial involving an investigational drug. Patients with a history of major surgery within 28 days were not allowed to participate Eligible patients had to have no known prior malignancies or known brain metastases; however, patients with adequately treated basal cell or squamous cell skin cancer, carcinoma in situ of the cervix, or any other cancer from which the patient had been disease-free for more than 5 years were permitted. No severe acute or chronic medical condition that may have impaired the ability to participate in the study or interfered with the interpretation of results was permitted. Adverse events from prior anti-cancer therapy were to have recovered to grade ≤ 1 (National Cancer Institute Common Terminology Criteria for Adverse Events [NCI-CTCAE] version 3.0) before randomisation. Patients were to have had no history of uncontrolled hypertension within 3 months before enrolment and no deep vein thrombosis within 4 weeks before randomisation. Patients were to have adequate organ and haematological function. Pregnant and breast-feeding women were excluded. Patients of reproductive potential were required to use effective methods of contraception

Participants

Patient characteristics and disease history (including prior anti-cancer treatments) appear to have been well balanced between the two treatment groups, as shown in Table 4.2. However, the timing of progression (i.e. whilst on oxaliplatin-containing therapy or the length of time after completion of

Superseded – see erratum

oxaliplatin-containing therapy) was not presented for patients in the aflibercept and placebo groups. The ERG requested this data from the manufacturer, which was provided; the proportion of patients progressing within 3 months (9% in the aflibercept group and 10.4% in the placebo group) and within 6 months (22.7% in the aflibercept group and 26.5% in the placebo group) of the start of first-line therapy was well balanced between the aflibercept and placebo groups. The ERG's clinical advisor stated that the proportion of patients progressing within 3 months and within 6 months of the start of first-line therapy appeared to reflect patients seen in practice.

The patient inclusion and exclusion criteria appear to have been generally appropriate; patients who had documented progression on or after completion of a single prior chemotherapeutic oxaliplatin-containing regimen for metastatic disease. Patients who experienced relapse on or within 6 months of completion of oxaliplatin-based adjuvant therapy were also eligible for inclusion. The MS contains a *post-hoc* subgroup analysis, excluding patients who received oxaliplatin as adjuvant therapy.

The MS does not state how many cycles of first-line oxaliplatin-containing palliative therapy participants in the trial had undergone. In current UK practice, patients who progress after a treatment break during intermittent first-line palliative chemotherapy are likely to be offered repeat treatment with the first-line chemotherapy regimen; the criteria for switching to second-line therapy is progression on treatment or symptomatic progression within six to eight weeks of completion of first-line chemotherapy. Therefore, the patients included in the trial were potentially at a less advanced stage of disease than those who would be offered second-line therapy in UK practice, who may undergo several cycles of first-line palliative therapy before progressing to second-line therapy.

In the UK between 2007 and 2009, an average 72% of colorectal cancer cases were diagnosed in people aged 65 years and over,[28] whilst in the trial only 33.5% of the aflibercept group and 38.9% of the placebo group were aged over 65 years. Therefore, the patients included in the trial were younger than those seen in standard UK practice. In addition, the ERG clinical advisor stated that the proportion of patients in the trial with an ECOG PS of 2 (2.2%) was lower than seen in other second-line clinical trials[29] and standard UK practice, 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. Each of these 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.

In clinical practice bevacizumab is often used in combination with FOLFOX as first-line chemotherapy for patients with mCRC; therefore, the proportion of patients who had undergone prior bevacizumab is likely to be higher in practice than in the trial.

Table 4.2 Baseline characteristics of patients in the VELOUR trial

Patient characteristics	**Aflibercept + FOLFIRI (n=612)**	**Placebo + FOLFIRI (n=614)**
Eastern Cooperative Oncology Group (ECOG) performance status (PS), n (%)		
0	349 (57.0)	350 (57.0)
1	250 (40.8)	250 (40.7)
2	13 (2.1)	14 (2.3)
Prior bevacizumab status (as provided at randomisation) n (%)		
Yes	186 (30.4)	187 (30.5)
No	426 (69.6)	427 (69.5)
Male, n (%)	365 (59.6)	353 (57.5)
Median age (range), years	61.0 (21-82)	61.0 (19-86)
Age > 65 years, %	33.5	38.9
Caucasian/white race, n (%)	548 (89.5)	523 (85.2)
Primary site, n (%)		
Colon	289 (47.2)	302 (49.2)
Recto-sigmoid	123 (20.1)	136 (22.1)
Rectum	197 (32.2)	174 (28.3)
Other (includes presumed colorectal primary, appendix, colon plus appendix, synchronous caecum, and rectum primary)	3 (0.5)	2 (0.3)
Median (range) time from first diagnosis to randomisation, months	14.62 (2.1-325.1)	13.67 (2.4-214.7)
Number of metastatic organs involved at baseline (excluding primary site), n (%)		
0	2 (0.3)	6 (1.0)
1	256 (41.8)	271 (44.1)
> 1	354 (57.8)	337 (54.9)
Metastatic organs involved at baseline (excluding primary site), n (%)		
Any site	610 (99.7)	608 (99.0)
Liver	459 (75.0)	431 (70.2)
Lung	271 (44.3)	277 (45.1)

Lymph nodes	173 (28.3)	181 (29.5)
Peritoneum	68 (11.1)	88 (14.3)
Liver metastasis, n (%)		
No liver metastasis, or liver and other metastases	459 (75.0)	468 (76.2)
Liver metastasis only	153 (25.0)	146 (23.8)
Prior chemotherapy, n (%)		
Adjuvant only (relapse within 6 months of completion of oxaliplatin-based adjuvant therapy)	60 (9.8)	64 (10.4)
Metastatic disease	450 (73.5)	442 (72.0)
Adjuvant and metastatic disease	102 (16.7)	108 (17.6)
Laboratory abnormalities at baseline		
Anaemia, %		
All grades	[redacted]	[redacted]
Grades 3-4	[redacted]	[redacted]
Elevated alkaline phosphatase, %		
All grades	[redacted]	[redacted]
Grades 3-4	[redacted]	[redacted]
Hypoalbuminaemia, %		
All grades	[redacted]	[redacted]
Grades 3-4	[redacted]	[redacted]
Elevated lactate dehydrogenase (> upper limit of normal), %	[redacted]	[redacted]

Intervention

Aflibercept was administered as an IV infusion over one hour on day one every two weeks (cycle) at a dose of 4mg/kg of body weight, followed immediately by FOLFIRI, administered as follows: irinotecan 180 mg/m^2 IV infusion over 90 minutes and leucovorin 400 mg/m^2 IV infusion over two hours at the same time on day one, followed by 5-FU 400 mg/m^2 IV bolus, followed by 5-FU 2,400 mg/m^2 continuous IV infusion over 46 hours at home. The treatment cycle is repeated every two weeks until disease progression or unacceptable toxicity. This FOLFIRI regimen is the standard FOLFIRI regimen used in the UK, therefore, the use of the intervention in the trial was directly relevant to the decision problem.

Dose adjustments for each treatment component individually and/or cycle delays (up to two weeks) were permitted in the event of toxicity. Premedication with atropine and antiemetics was permitted. Granulocyte colony-stimulating factor was used according to American Society of Clinical Oncology (ASCO) guidelines.

[redacted]The mean dose received was only slightly lower than the planned dose; the ERG considers that these dose reductions are likely to reflect the use of aflibercept plus FOLFIRI in clinical practice.

Comparator

A matching placebo was administered as an IV infusion over one hour on day one every two weeks, followed immediately by FOLFIRI (as described above).

[redacted]The mean dose received was similar to that received by the aflibercept group and only slightly lower than the planned dose; the ERG considers that these dose reductions are likely to reflect the use of FOLFIRI in clinical practice.

Outcomes

Overall survival

The primary outcome was OS, defined as the time interval from randomisation to death from any cause. Whilst OS is an objective and clinically relevant outcome it is subject to confounding from further treatments administered post-progression or withdrawal from study therapy for whatever reason. It was determined that 863 events were required to detect, with 90% power, a 20% risk reduction in the aflibercept group relative to the control group, using a log-rank test at an overall two-sided significance level of 0.0499, taking into account stopping boundaries for two interim analyses. The cut-off date for OS was the date of the 863rd event (7 February 2011).

Time-to-event parameters were estimated using Kaplan-Meier analysis. Primary comparisons between treatment groups were performed using stratified log-rank tests with stratification factors specified at randomisation. In pre-specified subgroup analyses, the consistency of the treatment effect on OS was evaluated with respect to stratification factors at randomisation (ECOG PS (0; 1; 2) and prior bevacizumab therapy), as well as a series of baseline characteristics: age (less than 65 years; 65 years or over), gender, location of primary tumour (colon; rectosigmoid; rectum; other), number of organs with metastases (more than 1; 1 or less), presence of liver metastases only (yes; no), prior history of hypertension, race (white; other), and geographical region (Western Europe; Eastern Europe; North America; South America; other). The hazard ratio (HR) and confidence interval (CI) estimates were provided using a Cox proportional hazards model.

The MS states that OS results for all pre-specified subgroups are shown in Figure B-8; however, Figure B8 only shows subgroup results for ECOG PS, prior bevacizumab therapy, prior hypertension, number of organs with metastases and presence of liver metastases. No subgroup assessment results were presented for age, gender, location of primary tumour, race or geographical region in the MS. Two subgroups were considered in more detail in the MS and included in the economic analysis; patients with liver metastases only and the subgroup of patients excluding those who relapsed on or within six months of adjuvant oxaliplatin treatment.

The manufacturer stated that the early and continued separation of the survival curves suggests that some patients experienced a sustained benefit following treatment with aflibercept plus FOLFIRI, which is not fully reflected by the median OS difference of 1.44 months. To justify this position the manufacturers presented the results of a piecewise model calculating HRs by 6-month periods. Whilst the HRs decreased over time the ERG found that results were still consistent with a proportional hazards assumption and hence with a constant survival benefit at all times. In addition, there were very few patients at risk at the later time points, and so the confidence intervals were very wide, indicating a lack of precision which reduces the reliability of the results.

Evaluation of the validity of proportional hazards assumptions was also performed using diagnostic plots (e.g. log cumulative hazard plots). The MS stated that the log-cumulative hazard plots demonstrated mild non-proportionality; the curves for the two treatment arms were non-parallel and crossed over one another (presented in Figure B-6 of the MS and Figure 4.2 of this report). However, because the curves for the two treatment arms were very close to being parallel, it is not certain that

Superseded – see erratum

hazards are non-proportional, therefore, this does not support their argument against using a proportional hazards approach to generating the survival curves, although this is a subjective judgement and there is uncertainty in interpreting the curves.

Based on their acceptance of non-proportionality and continued separation of the survival curves, the manufacturer estimated the mean OS in place of median OS. However, the suggested continued separation of the survival curves indefinitely is highly uncertain, given that data were unavailable for more than 36 months follow-up, and particularly when the OS curves are compared with the PFS curves, which converge at around 12 months (although the relationship between OS and PFS is uncertain).

By using mean survival, rather than median survival, the few patients with long term survival have greater weight in the analysis, so mean survival times are greater than median survival times. The survival curves (presented in Figure 4.1) do not suggest that median survival is inappropriate to use in the economic analysis.

The manufacturer stated that the mean OS could not be calculated from the trial data, as some patients were still alive at the end of study follow-up. The cut-off criteria used in the trial (the date of the 863rd event; 7 February 2011) meant that some patients had been enrolled in the study for less than a year (trial enrolment was from November 2007 to March 2010); therefore, follow-up for some of the trial patients was inadequate. The trial would have benefited from a minimum follow-up period for each patient enrolled, e.g. 24 months, ensuring a much more complete data set than the one presented, and removing the need to estimate mean OS.

In addition, the ERG does not entirely agree with the methods used to estimate mean OS. The manufacturer estimated the mean OS by fitting parametric functions to the trial data and extrapolating to provide complete curves. The following parametric distributions were tested; exponential, Weibull, log-logistic, log-normal and Gompertz.

The mean OS benefit estimate varies considerably, (from 3 months to) depending on which distribution is used, which indicates that the mean OS results presented may not be reliable. The ERG requested mean survival data for those patients who had died at the time of the most recent data analysis. The restricted mean OS until last death in each arm (area under Kaplan Meier curves) was provided by the manufacturer and the difference in OS between treatment groups was smaller

than in the manufacturer's estimations ([redacted] compared to 4.7 months). In addition, in the manufacturer's models the mean OS was estimated based on extrapolating survival curves to 15 years, which the ERG considers to be very long for this patient population and of uncertain validity given that the data extend to only 3 years. The ERG requested the estimates with the analysis truncated at 5 years; since it is unlikely that the treatment benefit would extend beyond 5 years. When the data are truncated at 5 years, the different models are more consistent, and the mean OS benefit is reduced (to [redacted] from 4.7 months).

Progression-free survival

Progression-free survival (PFS) was a secondary endpoint, defined as the interval from randomisation to the first observation of disease progression using Response Evaluation Criteria in Solid Tumors (RECIST) criteria (according to independent review committee (IRC) review) or death from any cause. Disease assessment was performed every 6 weeks until documented progression. For 42 patients who died prior to the implementation of the IRC review or who refused consent for this review, the investigator's tumour assessment was used.

Whilst IRC assessment of progression is more objective than an investigator's assessment in that it cannot be unblinded by severe or unusual adverse effects or influenced by the patient's status, it may miss symptom rather than tumour growth driven disease progression. As such it does not reflect assessment of disease in clinical practice. The discontinuation of treatment based on radiographic progression alone may have an impact on duration of treatment, and therefore costs. Clinical progression was included in a sensitivity analysis [redacted].

PFS is an important endpoint for the assessment of the activity of an individual agent in that, although it is not necessarily associated with improved overall survival, unlike directly measured overall survival, it is not subject to confounding by post-progression treatment.

Pre-specified subgroup analyses, as described above, were also performed to evaluate the consistency of the treatment effect on PFS.

Response rate

Objective response (complete response and partial response) was a secondary endpoint, assessed by the IRC, using RECIST criteria version 1.0. Disease assessment was performed every 6 weeks. Response was assessed in the subset of patients who had measureable disease at baseline; 531 patients in the aflibercept arm and 530 patients in the placebo arm. Response rate was compared between arms using a Cochran-Mantel-Haenszel test.

Adverse events

Adverse events were described, and the most frequent adverse events, anti-VEGF-associated events and most frequent biologic abnormalities were tabulated for the safety population. The safety population consisted of 611 patients in the aflibercept group who had received at least one dose of aflibercept (including 4 patients allocated to the placebo group who received at least one dose of aflibercept due to dosing irregularities), and 605 patients in the placebo group.

A meta-analysis of aflibercept adverse events was presented, including the VELOUR trial, a trial of patients with non-small cell lung cancer (the VITAL trial) and a trial of patients with metastatic pancreatic cancer (the VANILLA trial). The ERG requested details of the methods of this meta-analysis but these were not provided by the manufacturer. However, the manufacturer acknowledged, in their response to the ERG's points for clarification, that the results of the VITAL and VANILLA trials are not directly relevant to the safety profile likely to be shown by aflibercept in combination with FOLFIRI when used in the second-line treatment of mCRC in the UK, as the VITAL and VANILLA trials were conducted in different patient populations and used different chemotherapy backbones.

Discontinuation rates

Patients who discontinued treatment were followed up for disease progression using imaging every six weeks if discontinuation occurred before documented progression. Patients were also followed up for survival status every two months once progression was documented, until death or until the study cut-off date (whichever came first).

Health related quality of life

HRQoL was not assessed in the VELOUR trial. The manufacturer stated that the ongoing single-arm ASQoP trial (described in Section 4.1.7) has a secondary objective of documenting HRQoL in the

included patients to capture utility values not collected during the VELOUR trial. It is expected to report results in the next 12 months.

4.2.2 Summary of the quality of the included trials

The VELOUR trial was a good quality double-blind randomised placebo-controlled trial; results of the quality assessment are presented in Table B7 of the MS. There was a low risk of bias for all key domains using the Cochrane Risk of Bias tool.[30] The sample size was adequate, an appropriate method of randomisation was used and participants' general baseline characteristics were reported, which were broadly similar between the two treatment groups.

Patients, care-givers and outcome assessors were blinded to treatment group and aflibercept and placebo were supplied in indistinguishable sealed vials to preserve concealment of treatment allocation. PFS was assessed by an independent review committee (for the majority of patients), which means that even if patients and care-givers were able to guess their treatment allocation due to symptoms or adverse events, outcome assessment was still blinded for this outcome. Standard RECIST criteria were used for assessing response, which is appropriate. However, progression was only assessed in terms of radiographic criteria, rather than symptomatic progression; this was discussed in Section 4.2.1.

A high proportion of patients discontinued treatment in the trial; 97% withdrew from the aflibercept group and 97% withdrew from the placebo group; therefore, there was not an imbalance in drop-outs between groups, although the reasons for discontinuation were different between treatment groups. Cross-over between treatment groups was not allowed. Efficacy analyses were conducted on an intention-to-treat basis. Pre-planned and *post-hoc* subgroup analyses were performed.

The inclusion and exclusion criteria for the trial were clearly described and were generally appropriate, although, as discussed in Section 4.2.1, the trial population may be fitter and at a less advanced stage of disease than patients who would be eligible for aflibercept plus FOLFIRI in standard UK NHS practice.

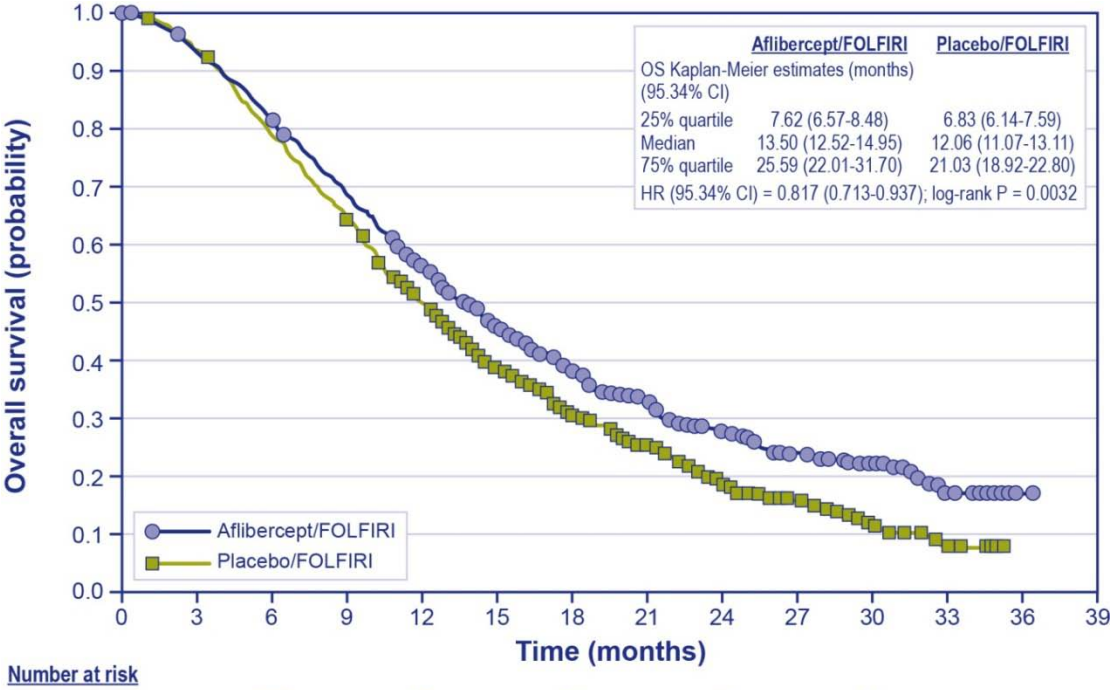
4.2.3 Summary of the results of the included trials

Overall survival

The cut-off date for the OS analysis was the date of the 863rd event (7 February 2011), with a median follow-up time of 22.28 months; 460 events (74.9%) were reported in the placebo group and 403 events (65.8%) were reported in the aflibercept group.

Figure 4.1 (Figure B-3 in the MS) presents the Kaplan-Meier plot for OS by treatment group. There was a statistically significant improvement in OS in the aflibercept group compared with the placebo group, with an estimated HR of 0.817 (95% CI 0.713 to 0.937; P=0.0032). The median OS was 13.5 months (95% CI 12.52 to 14.95) in the aflibercept group and 12.06 months (95% CI 11.07 to 13.11) in the placebo group; a difference of 1.44 months.

Figure 4.1 Kaplan-Meier estimates of overall survival



Number at risk						
Aflibercept:	612	498	311	148	75	33
Placebo:	614	485	286	131	51	14

Survival probability (%)					
Aflibercept:	81.9	56.1	38.5	28.0	22.3
Placebo:	79.1	50.3	30.9	18.7	12.0

The probabilities of surviving after 6 months, 12 months, 18 months, 24 months and 30 months were higher in the aflibercept group than the placebo group, as shown below the Kaplan-Meier plot in

Superseded – see erratum

Figure 4.1. However, the number of patients at risk at the later time points was very low (around 10% at 24 months and around 5% at 30 months); many patients enrolled in the study had not reached the later time points at the time of data cut-off, which reduces the reliability of the results at the later time points. Interpretation of the results beyond 30 months, where the numbers at risk were very small 10/6 (aflibercept/placebo) at 33 months and 1/0 at 36 months (and not even provided by the manufacturer, but taken from FDA statistical report[31]), is particularly difficult and subject to huge uncertainty.

A piecewise model calculating HRs by 6-month periods was presented in Table B9 of the MS, shown below as Table 4.3. The time periods used to calculate hazard ratios in this table were unequal, as all time points beyond 18 months were combined, making the HR unreliable for this time point. Therefore, the ERG asked the manufacturer to provide the OS HR for the following time intervals: 18 < t ≤ 24 months, 24 < t ≤ 30 months, 30 < t ≤ 36 months, and report the number of patients censored per treatment arm for each interval, as well as the number of patients at risk at t=36 months. The additional data provided by the manufacturer have been added to Table 4.3. There are very wide confidence intervals for each of the hazard ratios, particularly at the later time points, indicating a lack of precision. In addition, there were very few patients at risk at the later time points, which reduces the reliability of the results.

Table 4.3 Overall survival (months), piecewise Cox model with pre-defined intervals

Parameter	Time (months)	HR (95.34% CI) vs. Placebo + FOLFIRI	Aflibercept			Placebo		
			Censored	Death	Total	Censored	Death	Total
OS	t ≤ 6	0.860 (0.664-1.114)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
OS	6 < t ≤ 12	0.838 (0.673-1.043)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
OS	12 < t ≤ 18	0.782 (0.582-1.050)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
OS	t > 18	0.676 (0.463-0.988)						
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	[redacted]		[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Total			[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

The manufacturer stated that the early and continued separation of the survival curves suggests that some patients experienced a sustained benefit following treatment with aflibercept plus FOLFIRI, which is not fully reflected by the median OS difference of 1.44 months. However, the ERG does not agree with this interpretation: while the hazard ratios presented in Table 4.3 are changing over time they remain consistent with an assumption of proportional hazards, although the limited number of patients at longer follow-up times makes the estimation of hazard ratios uncertain, with wide confidence intervals.

The MS stated that the log-cumulative hazard plots demonstrated mild non-proportionality; the curves for the two treatment arms were non-parallel and crossed over one another (presented in Figure B-6 of the MS and Figure 4.2 below). However, because the curves for the two treatment arms were very close to being parallel, it is not certain that hazards are non-proportional, therefore, this does not support their argument against using a proportional hazards approach to generating the survival curves, although this is a subjective judgement and there is uncertainty in interpreting the curves. 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.

[redacted]

Superseded – see erratum

The manufacturer stated that the mean OS could not be calculated from the trial data, as some patients were still alive at the end of study follow-up. Therefore, they estimated the mean OS by fitting parametric functions to the trial data and extrapolating to provide complete curves. The following parametric distributions were tested; exponential, Weibull, log-logistic, log-normal and Gompertz. Evaluation of fit was performed using diagnostic plots (e.g. log cumulative hazard plots), Akaike information criterion (AIC) and Bayesian information criterion (BIC) tests, visual inspection of fit to the Kaplan-Meier data, and consideration of the biological and clinical plausibility of the extrapolation beyond the end of trial follow-up. Results were presented in Table B10 of the MS, also shown below as Table 4.4. Visual inspection plots were also presented in the MS, as Figures B-4 and B-5.

Table 4.4 AIC and BIC test results for overall survival functions

Function	Aflibercept + FOLFIRI		Placebo + FOLFIRI		Both arms (proportional hazards)	
	AIC	BIC	AIC	BIC	AIC	BIC
Weibull	1497.06	1505.89	1469.34	1478.18	2967.33	2982.67
Log-normal	1491.14	1499.98	1469.08	1477.92	2965.32	2980.66
Log-logistic	1479.70	1488.53	1457.19	1466.03	2939.84	2955.17
Exponential	1535.40	1539.82	1549.73	1554.15	3085.14	3095.36
Gompertz	1524.47	1533.30	1510.60	1519.44	3037.04	3052.38

AIC, Akaike information criterion; BIC, Bayesian information criterion
Source: Joulain *et al*. 2012

Table B11 of the MS (presented below as Table 4.5) presents the estimated mean OS based on an extrapolation to 15 years follow-up, using the exponential, Weibull, log-logistic and log-normal distributions. The mean OS estimate varies considerably, depending on which distribution is used, which indicates that the mean OS results presented may not be reliable. The ERG requested mean survival data for those patients who had died at the time of the most recent data analysis. The restricted mean OS until last death in each arm (area under Kaplan Meier curves) was provided by the manufacturer[redacted]. The ERG acknowledges that these figures are likely to be an

underestimate, as some patients with long survival times will have been censored. These estimates are, however, based on the actual data, rather than an extrapolated model. It is unclear whether the difference in mean OS between the aflibercept group and placebo group based only on the data is more accurate than the highly variable, and considerably higher, estimates from the models, presented in Table 4.5 below. Table 4.5 Estimated mean overall survival at 15 years

	Mean overall survival at 15 years (months)		
	Aflibercept + FOLFIRI	**Placebo + FOLFIRI**	**Difference**
Log-logistic distribution (by treatment separately[a])	**22.8**	**18.1**	**4.7**
Weibull distribution (by treatment separately)	17.9	14.9	3.0
Log-normal distribution (by treatment separately)	[redacted]	[redacted]	[redacted]
Exponential distribution (by treatment separately)	[redacted]	[redacted]	[redacted]
[a] Function providing best fit.			

These estimates of mean OS do not allow for the possibility that the treatment benefit of aflibercept would reduce to zero before 15 years. The ERG requested recalculation of these estimates with the analysis truncated at 5 years and 10 years, since 15 years seems very long for this patient population, who have a median overall survival of around 10.5 to 18 months. The mean OS data for the different time horizons, for different parametric distributions, are presented below as Table 4.6. When the data are truncated at 5 years, the different models are more consistent[redacted]
.

Table 4.6 Estimated mean overall survival at different time horizons

Overall Survival	Placebo/FOLFIRI (N=614)	Aflibercept/ FOLFIRI (N=612)	Comparison
Number of patients	614 (100%)	612 (100%)	
Number of events	460 (74.9%)	403 (65.8%)	
Median in months* [95% CI]	12.06 [11.07 ; 13.08]	13.50 [12.52 ; 14.95]	Median difference = 1.44
Median in months assuming LogLogistic (a)	[redacted]	[redacted]	[redacted]
Restricted mean over 5 years in months assuming LogLogistic (a)	[redacted]	[redacted]	[redacted]
Restricted mean over 10 years in months assuming LogLogistic (a)	[redacted]	[redacted]	[redacted]
Restricted mean over 15 years in months assuming LogLogistic (a)	18.1	22.8	Mean difference = 4.7
Median in months assuming Weibull (a)	[redacted]	[redacted]	[redacted]
Restricted mean over 5 years in months assuming Weibull (a)	[redacted]	[redacted]	[redacted]
Restricted mean over 10 years in months assuming Weibull (a)	[redacted]	[redacted]	[redacted]
Restricted mean over 15 years in months assuming Weibull (a)	14.9	17.9	Mean difference =3.0
Median in months assuming Lognormal (a)	[redacted]	[redacted]	[redacted]
Restricted mean over 5 years in months assuming Lognormal (a)	[redacted]	[redacted]	[redacted]
Restricted mean over 10 years in months assuming Lognormal (a)	[redacted]	[redacted]	[redacted]
Restricted mean over 15 years in months assuming Lognormal (a)	[redacted]	[redacted]	[redacted]
Median in months assuming Exponential (a)	[redacted]	[redacted]	[redacted]
Restricted mean over 5 years in months assuming Exponential (a)	[redacted]	[redacted]	[redacted]
Restricted mean over 10 years in months assuming Exponential (a)	[redacted]	[redacted]	[redacted]
Restricted mean over 15 years in months assuming Exponential (a)	[redacted]	[redacted]	[redacted]
Median in months assuming Gompertz (a)	[redacted]	[redacted]	[redacted]
Restricted mean over 5 years in months assuming Gompertz (a)	[redacted]	[redacted]	[redacted]
Restricted mean over 10 years in months assuming Gompertz (a)	[redacted]	[redacted]	[redacted]
Restricted mean over 15 years in months assuming Gompertz (a)	[redacted]	[redacted]	[redacted]
* Refer to Kaplan-Meier curve (a) By treatment group independently			

The manufacturer stated that they consider 15 years to be an appropriate time horizon; while only a minority of patients with mCRC will survive for longer than 5 years, it is important that the time horizon is sufficient to capture survival for that minority of patients who do survive for extended periods of time. However, the ERG considers that it is unlikely that the treatment benefit would extend beyond 5 years, and it may well be shorter.

Superseded – see erratum

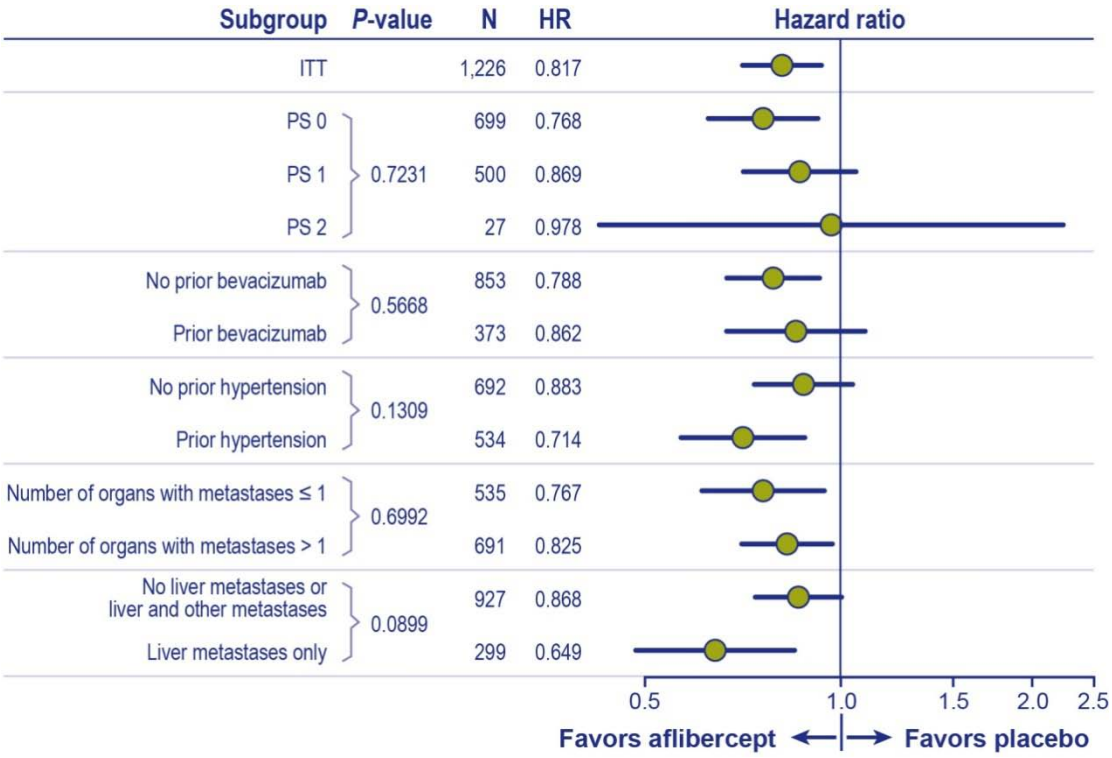
By using mean survival, rather than median survival, the few patients with long term survival have greater weight in the analysis. The survival curves presented in Figure 4.1 do not suggest that median survival is inappropriate to use in the economic analysis.

Overall survival subgroup analyses

As demonstrated in Figure 4.3 below (Figure B-8 in the MS) there was no evidence of interaction between treatment groups and the following stratification factors; ECOG PS, prior bevacizumab therapy. Similarly, there was no evidence of interaction between treatment groups and the following patient characteristics; prior hypertension, number of organs with metastases, indicating an overall consistent treatment effect across these subgroups. However, there was a significant interaction at the 10% level (P=0.0899) indicating that patients with liver metastases only had a greater OS benefit, compared with patients with no liver metastases or liver and other metastases. For the subgroup of patients with liver metastases only, median OS [redacted] HR 0.649 (95% CI 0.49 to 0.86). [redacted]

Whilst there was no evidence of a significant interaction between treatment groups for most of the baseline patient characteristics, the results suggest that patients with less advanced disease (ECOG PS 0, number of organs with metastases ≤1, liver only metastases) may be more likely to benefit from aflibercept. 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 VELOUR trial.

Figure 4.3 Overall survival by pre-specified subgroups



No subgroup assessment results were presented in the MS for the following pre-specified subgroups: age, gender, location of primary tumour, race or geographical region. However, Figure 7 of the EMA assessment report presents HRs for the subgroups age, gender, race and geographical region; no significant treatment interaction was noted for any of these participant characteristics.[32]

A *post-hoc* subgroup analysis, excluding 124 patients who relapsed on or within 6 months of oxaliplatin-based adjuvant therapy, demonstrated a non-statistically significant greater treatment effect; median OS 13.8 months (95% CI 12.68 to 15.44) in the aflibercept group and 11.9 months (95% CI 10.88 to 13.01) in the placebo group; unadjusted HR 0.78 (95% CI 0.68 to 0.90); adjusted HR 0.80 (95% CI 0.69 to 0.92), compared with patients who only received oxaliplatin as adjuvant therapy; unadjusted HR 1.09 (95% CI 0.70 to 1.69); adjusted HR 1.05 (95% CI 0.66 to 1.65); P value for interaction with treatment 0.1265.

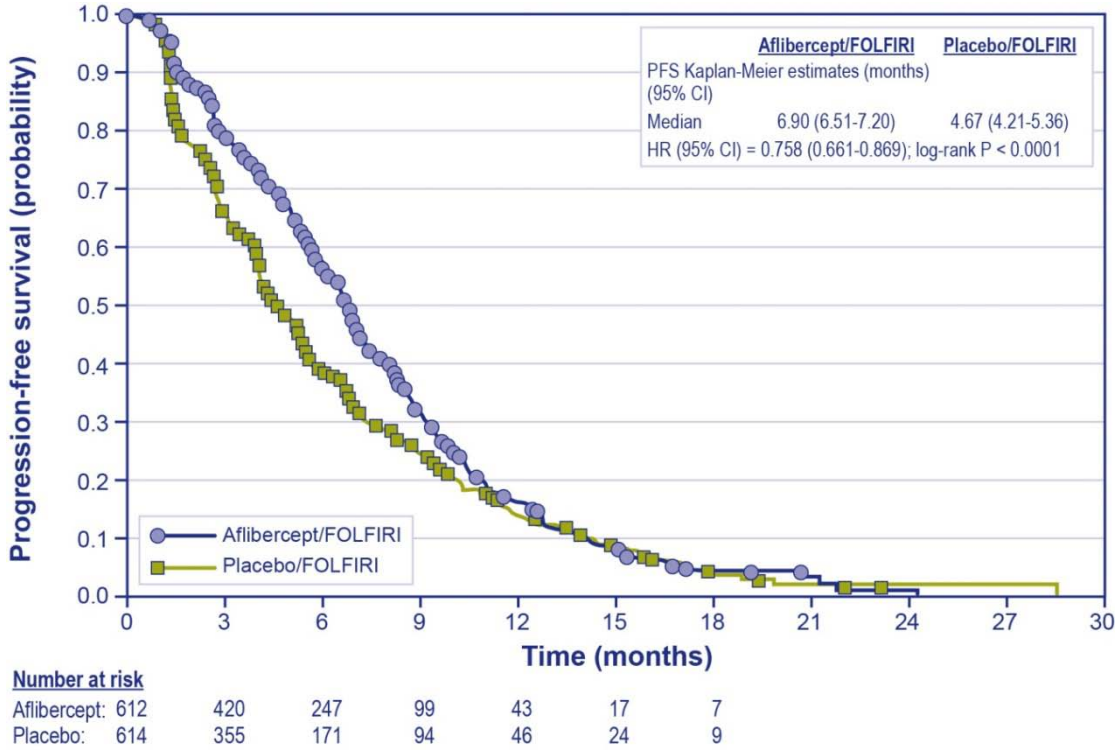
[redacted]

[redacted]

Progression-free survival

Figure 4.4 (Figure B-7 in the MS) presents the Kaplan-Meier plot for PFS by treatment group. There was a statistically significant improvement in PFS for the aflibercept group compared with the placebo group, with an estimated HR of 0.758 (95% CI 0.661 to 0.869; P=0.0001). The median PFS was 6.90 months (95% CI 6.51 to 7.20) for the aflibercept group and 4.67 months (95% CI 4.21 to 5.36) for the placebo group; a difference of 2.23 months.

Figure 4.4 Kaplan-Meier estimates of progression free survival



As shown below the Kaplan-Meier plot, the number of patients at risk at the later time points was very low (less than 35% at 6 months, around 15% at 9 months, around 7% at 12 months and around 3% at 15 months).

[redacted]

[redacted] As discussed earlier, the investigators' own assessment is more reflective of how progression is assessed in clinical practice. The FDA Medical Review states that there was a high rate of discrepancy between the investigator assessments and the IRC assessments; 46% in the placebo group and 39% in the aflibercept group (Table 35 of the FDA Medical Review).[33]

Progression-free survival subgroup analyses

Consistent with the OS subgroup analyses, there was no evidence of interaction between treatment groups and the following stratification factors; ECOG PS, prior bevacizumab therapy. Similarly, there was no evidence of interaction between treatment groups and the following demographic characteristics; prior hypertension, number of organs with metastases, indicating an overall consistent treatment effect across these subgroups. However, consistent with the OS subgroup analysis, there was a significant interaction at the 10% level [redacted] indicating that patients with liver metastases only had a greater PFS benefit, compared with patients with no liver metastases or liver and other metastases.

[redacted]

No subgroup assessment results were presented in the MS for the following pre-specified subgroups: age, gender, location of primary tumour, race or geographical region.

[redacted]

Superseded – see erratum

[redacted]

Response rate

Table 4.7 (Table B12 in the MS) presents the response rates for the evaluable population (1061/1226 patients), as assessed by the IRC. There was a statistically significant improvement in response rate in the aflibercept group compared with the placebo group; 19.8% (95% CI 16.4 to 23.2) in the aflibercept group and 11.1% (95% CI 8.5 to 13.8) in the placebo group.

Table 4.7 Response rates

Response rates (secondary endpoint, evaluable population)	Aflibercept + FOLFIRI (n = 612)	Placebo + FOLFIRI (n = 614)	*P* value
Evaluable population, n	531	530	
Response rate (complete or partial response), n (%) [95% CI, %]	105 (19.8) [16.4-23.2]	59 (11.1) [8.5-13.8]	< 0.001
Complete response, n (%)	0	2 (0.4)	—
Partial response, n (%)	105 (19.8)	57 (10.8)	—
Stable disease, n (%)	350 (65.9)	344 (64.9)	—
Progressive disease, n (%)	55 (10.4)	114 (21.5)	—
Not evaluable, n (%)	21 (4.0)	13 (2.5)	

The evaluable population included all randomised patients with measurable disease at study entry, with at least one valid post-baseline tumour evaluation, and who gave consent for third-party review.

Source: Van Cutsem *et al*., 2012.

Adverse events

Treatment-emergent adverse events were reported in 99.2% of aflibercept patients and 97.9% of placebo patients. Grade 3-4 treatment-emergent adverse events were reported in 83.5% of aflibercept patients and 62.5% placebo patients. The majority of grade 3-4 adverse events occurred during early treatment cycles.

Table B15 of the MS (presented as Table 4.8 below) presents the most frequent adverse events (incidence ≥20% or ≥5% higher in the aflibercept arm), other anti-VEGF-associated events, and most frequent biologic abnormalities. 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; particularly grade 3 diarrhoea, grade 3 hypertension, and all grades of: stomatitis and ulceration, infections and infestations, hypertension, haemorrhage, epistaxis, dysphonia, headache, proteinuria.

[redacted]

The MS also presented a meta-analysis including the VELOUR trial and two additional trials; VITAL which assessed 6 mg/kg aflibercept plus docetaxel for patients with non-small cell lung cancer, and VANILLA which assessed 4 mg/kg aflibercept plus gemcitabine for patients with metastatic pancreatic cancer. The results of the meta-analysis were similar to the results of the VELOUR trial presented above. The meta-analysis of grade 3-4 anti-VEGF adverse events found a statistically significant higher risk of hypertension, proteinuria and haemorrhage in the aflibercept group than the placebo group. However, the manufacturer acknowledged, in their response to the ERG's points for clarification, that the results of the VITAL and VANILLA trials are not directly relevant to the safety profile likely to be shown by aflibercept in combination with FOLFIRI when used in the second-line treatment of mCRC in the UK, as the VITAL and VANILLA trials were conducted in different patient populations and used different chemotherapy backbones.

Table 4.8 Most frequent adverse events (events (incidence ≥20% or ≥5% higher in the aflibercept arm), other anti-VEGF-associated events, and most frequent biologic abnormalities: safety population

Adverse event[a]	Aflibercept + FOLFIRI (n = 611)			Placebo + FOLFIRI (n = 605)		
	All grades (%)	Grade 3 (%)	Grade 4 (%)	All grades (%)	Grade 3 (%)	Grade 4 (%)
Any	99.2	62.0	21.4	97.9	45.1	17.4
Diarrhoea (PT)	69.2	19.0	0.3	56.5	7.6	0.2
Asthenic conditions (HLT)	60.4	16.0	0.8	50.2	10.4	0.2
Stomatitis and ulceration (LT)	54.8	13.6	0.2	34.9	5.0	—
Nausea (PT)	53.4	1.8	—	54.0	3.0	—
Infections and infestations (SOC)	46.2	11.0	1.3	32.7	6.1	0.8
Hypertension	41.4	19.1	0.2	10.7	1.5	—
Haemorrhage	37.8	2.8	0.2	19.0	1.7	—
Epistaxis	27.7	0.2	—	7.4	—	—
GI and abdominal pains (HLT)	34.0	5.1	0.3	29.1	3.1	0.2
Vomiting (PT)	32.9	2.6	0.2	33.4	3.5	—
Decreased appetite (PT)	31.9	3.4	—	23.8	1.7	0.2
Weight decreased	31.9	2.6	—	14.4	0.8	—
Alopecia (PT)	26.8	—	—	30.1	—	—
Dysphonia (PT)	25.4	0.5	—	3.3	—	—
Constipation (PT)	22.4	0.8	—	24.6	1.0	—
Headache (PT)	22.3	1.6	—	8.8	0.3	—

Adverse event[a]	Aflibercept + FOLFIRI (n = 611)			Placebo + FOLFIRI (n = 605)		
	All grades (%)	Grade 3 (%)	Grade 4 (%)	All grades (%)	Grade 3 (%)	Grade 4 (%)
Palmar-plantar erythrodysesthesia syndrome	11.0	2.8	—	4.3	0.5	—
Other anti-VEGF–associated events						
Arterial thromboembolic events	2.6	0.8	1.0	1.5	0.5	—
Venous thromboembolic events	9.3	3.1	4.7	7.3	2.6	3.6
Fistula from GI origin	1.1	0.3	—	0.3	0.2	—
Fistula from other than GI origin	0.3	—	—	0.2	—	—
GI perforation	0.5	0.2	0.3	0.5	0.2	0.2
Biologic abnormalities						
Haematologic						
Anaemia	82.3	3.3	0.5	91.1	3.5	0.8
Neutropaenia	67.8	23.1	13.6	56.3	19.1	10.4
Neutropaenic complications	6.5	4.4	1.3	3.0	1.7	1.2
Thrombocytopaenia	47.7	1.7	1.7	33.8	0.8	0.8
Non-haematologic						
Proteinuria	62.2	7.5	0.3	40.7	1.2	—
ALT increased	47.3	2.5	0.2	37.1	2.2	—

ALT, alanine aminotransferase; GI, gastrointestinal; HLT, high-level term; PT, preferred term; SOC, system organ class; VEGF, vascular endothelial growth factor.

[a] Grades were determined according to the National Cancer Institute Common Terminology Criteria of Adverse Events, version 3.0.

Source: Van Cutsem *et al*., 2012.

Discontinuation rates

At the time of data cut-off 598/614 (97.4%) patients in the aflibercept group and 593/612 (96.9%) patients in the placebo group had discontinued treatment. Participant flow is presented in Figure B-2 of the MS. Treatment-emergent adverse events led to permanent discontinuation of treatment in 26.8% patients in the aflibercept group and 12.1% patients in the placebo group. The adverse events (all grades) leading most frequently to permanent discontinuation of study treatment were asthenic conditions (3.8% versus 1.3%), infections (3.4% versus 1.7%), diarrhoea (2.3% versus 0.7%) and hypertension (2.3% versus 0%).[redacted]

The ERG asked the manufacturer to provide data on the duration, dose and number of administrations of second-line treatment. The manufacturer provided the data requested.

[redacted]

Superseded – see erratum

Table 4.9 All further anti-cancer therapy

	Placebo/Folfiri (N=614)	Aflibercept/Folfiri (N=612)
Any further therapy	[redacted]	[redacted]
Surgery	[redacted]	[redacted]
Radiotherapy	[redacted]	[redacted]
Systemic anti-cancer treatment	[redacted]	[redacted]
Biologics / Small molecules	[redacted]	[redacted]
Cetuximab	[redacted]	[redacted]
Bevacizumab	[redacted]	[redacted]
Panitumumab	[redacted]	[redacted]
Other	[redacted]	[redacted]
Chemotherapy	[redacted]	[redacted]
Fluoropyrimidine	[redacted]	[redacted]
Irinotecan	[redacted]	[redacted]
Other	[redacted]	[redacted]
Oxaliplatin	[redacted]	[redacted]
Other[a]	[redacted]	[redacted]

4.3 Conclusions of the clinical effectiveness section

The MS presented a well conducted systematic review of aflibercept for the second-line treatment of mCRC, which identified one double-blind randomised placebo-controlled trial; the VELOUR trial. The search strategy was adequate, inclusion screening and data extraction were undertaken in duplicate to reduce error and bias, adequate details of the included trial were presented, including an appropriate quality assessment; the trial was good quality.

The VELOUR trial was conducted in 28 countries; only 99 of the 1,226 participants were from the UK. 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 practice. 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.

The trial found a small, but statistically significant, increase in OS for patients in the aflibercept group compared with patients in the placebo group; the difference in median OS was 1.44 months. The survival increase was sustained, with a 22.3% survival probability in the aflibercept group compared with 12% in the placebo group at 30 months. However, there were very small numbers of patients

included in the analysis at the later time points, which reduces the reliability of the longer-term results. The ERG considers that the suggested continued separation of the survival curves indefinitely is highly uncertain, given that data were unavailable for more than 36 months follow-up, and particularly when the OS curves are compared with the PFS curves, which converge at around 12 months. The ERG considers that it is unlikely that the treatment benefit could extend beyond 5 years, and it may well be shorter.

There was also a statistically significant increase in PFS for the aflibercept group; the difference in median PFS between treatment groups was 2.23 months. Response rate was also significantly improved in the aflibercept group compared with the placebo group (19.8% compared with 11.1%). The adverse effects profile was significantly worse for aflibercept plus FOLFIRI than FOLFIRI alone. Grade 3-4 adverse events were more frequent in the aflibercept group than the placebo group (84% compared with 63%), serious adverse events were also more common in the aflibercept group (49% compared with 33%); 13 patients in the aflibercept group and 6 patients in the placebo group died as a result of adverse events. Treatment-emergent adverse events led to permanent discontinuation of treatment in 26.8% patients in the aflibercept group and 12.1% patients in the placebo group. HRQoL was not assessed in the VELOUR trial.

The MS included an estimation of mean OS benefit based on an extrapolation of the data, which was considerably longer than the median OS benefit reported, and also the restricted mean OS data supplied by the manufacturer upon request. The ERG considers that the estimates of mean OS may not be reliable; the estimate varies considerably depending on which distribution is used. Truncating the data at 5 years, rather than 15 years, improved consistency across the different models, and reduced the OS benefit (data supplied by manufacturer upon request).

In summary, there appears to be a statistically significant, but clinically small, increase in OS and PFS with aflibercept plus FOLFIRI, compared with FOLFIRI alone. However, grade 3-4 adverse events are common, leading to discontinuation of treatment in a large number of patients. There are currently no data on HRQoL for patients taking aflibercept.

4.4 Other relevant evidence for consideration

4.4.1 Evidence for other anti-VEGF therapy (bevacizumab)

As discussed in Section 3.3, a formal indirect comparison between aflibercept and bevacizumab is not considered appropriate. However, as bevacizumab is the only other licenced biological therapy that

works in the same way as aflibercept, the efficacy and adverse effects of bevacizumab in the second-line treatment of mCRC are presented below; efficacy results are also presented in Table 4.10.

The E3200 trial assessed the safety and efficacy of FOLFOX-4 + bevacizumab versus FOLFOX-4 alone versus bevacizumab alone in patients with advanced or metastatic CRC previously treated with irinotecan and 5-fluorouracil.[15] Median OS was significantly longer for patients in the FOLFOX-4 + bevacizumab arm (12.9 months) compared with patients in the FOLFOX-4 alone arm (10.8 months) (HR=0.75, p=0.0011). Median PFS was also significantly longer for patients in the FOLFOX + bevacizumab arm (7.3 months) compared with patients in the FOLFOX-4 alone arm (4.7 months) (HR=0.61, p<0.0001). Grade 3-4 adverse events were more common with FOLFOX-4 + bevacizumab arm than the FOLFOX-4 alone arm; 75% versus 61%; rates of grade 3-4 neuropathy, hypertension, bleeding and vomiting were higher in patients receiving bevacizumab. There was no significant difference in the incidence of adverse events leading to treatment discontinuation (23.4% versus 23.9%).

The ML18147 trial assessed bevacizumab + fluoropyrimidine-based chemotherapy versus fluoropyrimidine-based chemotherapy alone in patients with mCRC with disease progression following first-line treatment with bevacizumab.[13] The choice between irinotecan or oxaliplatin alongside fluoropyrimidine-based chemotherapy was switched depending on first-line usage of either oxaliplatin or irinotecan. Median OS was significantly longer for patients in the bevacizumab + fluoropyrimidine-based chemotherapy arm (11.2 months) compared with patients in the fluoropyrimidine-based chemotherapy alone arm (9.8 months) (HR=0.81, p=0.0062). Median PFS was also significantly longer for patients in the bevacizumab + fluoropyrimidine-based chemotherapy arm (5.7 months) compared with patients in the fluoropyrimidine-based chemotherapy alone arm (4.1 months) (HR=0.68, p<0.0001). Grade 3-5 adverse events were more common in the bevacizumab + fluoropyrimidine-based chemotherapy arm than the fluoropyrimidine-based chemotherapy alone arm; 64% versus 57%; rates of grade 3-5 bleeding or haemorrhage, gastrointestinal perforation and venous thromboembolisms were higher in patients receiving bevacizumab. More patients discontinued treatment due to adverse events in the bevacizumab + fluoropyrimidine-based chemotherapy arm (16% versus 9%).

A meta-analysis of bevacizumab + FOLFIRI as second-line therapy in mCRC patients pre-treated with oxaliplatin, which included no randomised comparisons with FOLFIRI alone (most were small single arm retrospective studies) generated a weighted median PFS of 8.3 months (range 3.9 to 11.6)

and weighted median OS of 17.2 months (range 10.9 to 22.4).[16] The results from this analysis of small uncontrolled studies has to be interpreted with caution; there is an extremely high risk of publication bias and it is highly probable that the median PFS and OS values are very optimistic.

Table 4.10 Efficacy of other anti-VEGF therapy (bevacizumab)

	Bevacizumab (E3200 trial)		Bevacizumab (ML18147 trial)*	
	Bevacizumab + FOLFOX4	FOLFOX4	Bevacizumab + chemotherapy	Chemotherapy
Median OS (months)	12.9	10.8	11.2 (95% CI: 10.4 to 12.2)	9.8 (95% CI: 8.9 to 10.7)
Median OS treatment difference (months)	2.1 HR = 0.75		1.4 HR = 0.81 (95% CI: 0.69 to 0.94)	
Median PFS (months)	7.3	4.7	5.7 (95% CI: 5.2 to 6.2)	4.1 (95% CI: 3.7 to 4.4)
Median PFS treatment difference (months)	2.6 HR = 0.61		1.6 HR = 0.68 (95% CI: 0.59 to 0.78)	

*In patients with mCRC with disease progression following first-line treatment with bevacizumab

4.4.2 Additional evidence of aflibercept for the treatment of mCRC

Given the limited data available regarding aflibercept in mCRC, the ERG thinks it is important to understand the possibilities of this drug across closely related indications. The VELOUR trial has provided evidence for its efficacy in the second-line setting; efficacy though statistically significant is clinically quite modest (increase in median PFS of 2.23 months). An additional trial has been conducted in the first-line setting: the AFFIRM trial. This was a randomised trial in which patients with mCRC were randomised to either aflibercept plus mFOLFOX6 or mFOLFOX6 alone. The only available publication of this trial (abstract from the 14th World Congress on Gastrointestinal Cancer O-0024) stated that the mFOLFOX6 study arm was included ‘as a calibrator’; the study was not powered for a statistical comparison.

It is unclear whether the study was planned to be underpowered. The ERG requested further details of this trial but the manufacturer declined to provide these stating that the AFFIRM trial is not directly relevant to the decision problem. They did state that publication is planned for late 2013.

The primary endpoint was PFS at 12 months. Patients with previously untreated mCRC, ECOG PS 0-2, and adequate organ function were randomized 1:1 to receive aflibercept 4 mg/kg plus mFOLFOX6 or mFOLFOX6, every 2 weeks, and were stratified by ECOG PS (0-1/2), prior adjuvant therapy (yes/no), and metastases confined to liver (yes/no). Patients were to be treated to disease progression. PFS and objective response were assessed according to RECIST criteria by an Independent Review Committee blinded to treatment allocation.

236 patients (median age, 62.5 years; male, 61%; PS 0-1, 97.5%; prior adjuvant therapy, 10.2%; liver metastases only, 28.4%) were randomized to aflibercept plus mFOLFOX6 (n = 119) or mFOLFOX6 alone (n = 117). Baseline characteristics were similar in both arms. PFS at 12 months was 25.8% (95% CI: 17.2 to 34.4) for aflibercept plus mFOLFOX6 and 21.2% (95% CI: 12.2 to 30.3) for mFOLFOX6 alone. Response rate was 49.1% (95% CI: 39.7-58.6) with aflibercept plus mFOLFOX6 and 45.9% (95% CI: 36.4-55.7) with mFOLFOX6 alone, and median PFS was 8.48 months (95% CI: 7.89-9.92) and 8.77 months (95% CI: 7.62-9.27) respectively. Overall survival data had limited maturity, with less than 43% death events in each arm. Grade 3-4 treatment-emergent adverse events occurred in 90.8% of the aflibercept plus mFOLFOX6 arm and 75.0% of the mFOLFOX6 arm. Treatment-emergent adverse events leading to death occurred in 6.7% of the aflibercept plus mFOLFOX6 arm and 1.7% of the mFOLFOX6 arm. Grade 3-4 adverse events with >5% higher incidence in the aflibercept plus mFOLFOX6 arm relative to the mFOLFOX6 arm were hypertension, proteinuria, neutropaenia and diarrhoea.

The authors concluded that in this phase 2 study, 12-month PFS with aflibercept plus mFOLFOX6 appeared to be similar to mFOLFOX6 alone, although the study was not powered for comparison. The safety profile of aflibercept was consistent with that seen in prior studies.

This trial would appear to demonstrate a failure of aflibercept to improve patient outcomes above those that can be achieved with mFOLFOX6 alone as first-line therapy for mCRC.

The ERG agree with the manufacturer that the use of this trial as evidence in the second-line setting may not be appropriate as agents with efficacy in a later setting do not always demonstrate this is earlier settings.

5 COST EFFECTIVENESS

The manufacturer's initial economic submission to NICE included:

1. A systematic review of the cost-effectiveness evidence for aflibercept in patients with mCRC;
2. A *de-novo* economic evaluation including an electronic version of the model in Microsoft Excel;
3. The protocols and results of two unpublished observational studies conducted by the manufacturer:
    a. Utility estimates in mCRC – the mCRC Utilities study;
    b. Resource use estimates in mCRC – the UK mCRC Resource Use study.

Following the points of clarification raised by the ERG, a number of addenda were submitted by the manufacturer. These included:

1. Additional information and results from the Key Opinion Leader (KOL) survey undertaken by the manufacturer with UK oncologists and used to inform resource use inputs (those related to the treatment of adverse events and medical management in a community setting);
2. Goodness of fit results for the extrapolation of PFS and OS using the Gompertz function.

This section of the ERG report provides a critical review of the quality of the economic evaluation; key assumptions, possible limitations and any remaining uncertainties are highlighted. These issues are subsequently investigated through additional exploratory analyses undertaken by the ERG in Section 6.

5.1 ERG comment on manufacturer's review of cost-effectiveness evidence

5.1.1 Searches

The manufacturer's submission described the search strategies used to identify cost-effectiveness studies relevant to this appraisal of aflibercept for the treatment of mCRC. The review was performed as part of a wider systematic review aimed to identify utility, resource use and cost estimates relevant to the appraisal. Search strategies were only briefly described in the main submission, however full details were provided in the Appendix 10 of the MS.

The electronic databases MEDLINE and MEDLINE In-Process (via PubMed), EMBASE(Dialog), EconLit (EBSCO), the Cochrane Library (Wiley) including the NHS Economic Evaluation Database (NHS EED) and the Health Technology Assessment (HTA) database were searched. In addition to this, abstracts of conference proceedings from American Society of Clinical Oncology (ASCO), European Society for Medical Oncology (ESMO), World Conference on Gastrointestinal Cancer, and International Society for Pharmacoeconomics and Outcomes Research (ISPOR) were searched. Regulatory organisation websites were also searched, including NICE, the Scottish Medicines Consortium (SMC), and the All Wales Medicines Strategy Group (AWMSG).

Database searches were performed on 27 June 2012. Internet searches were performed between 27 June and 6 July 2012. Search strategies for each database were documented in Tables 10-8 to 10-12 of the Appendix 10 of the MS. The searches covered the period 2010 to 27 June 2012. Searches were not performed for publications before 2010 because the NICE multiple-technology appraisal for cetuximab, bevacizumab, and panitumumab was published that year and provided relevant economic literature to that point. Further, the manufacturer did not expect there to be any aflibercept economic analyses prior to 2010. The searches were limited to English language publications, and excluded animal studies as well as the publication types ‘review’ and ‘case reports’, and the MeSH term ‘Review Literature as Topic’.

The searches were appropriate and comprehensive, and included the use of both subject indexing terms and free text searching. Field searching, Boolean operators and truncation were used where required. All NICE required databases were searched, as well as abstracts of conference proceedings and regulatory body websites. An update of the searches closer to submission might have been useful.

Methodological search filters were included to identify economic studies and utilities in MEDLINE, EMBASE and the Cochrane Library. The economic study design search filter may have excluded potentially useful records from the NHS EED database in the Cochrane Library search as this database is already limited by study design.

5.1.2 Inclusion/exclusion criteria used for study selection

The manufacturer undertook a series of systematic searches to identify published economic evaluations, including cost-effectiveness, and cost-utility analyses, as well as utility, resource use and

cost studies. The ERG believes that the inclusion/exclusion criteria were appropriate, and would have identified any relevant studies.

5.1.3 Studies included and excluded in the cost effectiveness review

The manufacturer's systematic review did not identify any studies that evaluated the cost-effectiveness of aflibercept for patients with mCRC. The ERG identified one potentially relevant cost-effectiveness study by Morlock *et al.*[35],2013, that was not included in the MS as it was published after the manufacturer conducted their search. The study aimed to assess the cost-effectiveness of bevacizumab + chemotherapy vs. aflibercept + chemotherapy in patients with mCRC previously treated with bevacizumab. It was not stated which treatments were included within chemotherapy. The study concluded that the bevacizumab + chemotherapy was less costly ($39,104 less per treated patient) and had similar effectiveness (OS 13.3 vs. 12.5 months; HR 0.94 (95% CI 0.70-1.26); 0.498 vs. 0.479 QALYs) to aflibercept + chemotherapy in second-line treatment of mCRC. The key cost-effectiveness drivers were identified as the price of aflibercept (which was higher than that of bevacizumab), and the higher rate of adverse events on the aflibercept + chemotherapy treatment. Since the study was presented as a meeting abstract, the ERG was unable to assess the quality of this cost-effectiveness analysis. Moreover, the study does not directly inform the cost-effectiveness of aflibercept using comparators relevant in the context of this appraisal.

5.1.4 Conclusions of the systematic review

There is a paucity of evidence on the cost-effectiveness of aflibercept. The only study[35], identified by the ERG, was conducted in the US and established a comparison against a treatment that was not considered a relevant comparator within this appraisal, and therefore of limited relevance. Furthermore, there is a lack of detail that precludes any formal quality assessment of the study.

5.2 Summary and critique of manufacturer's submitted economic evaluation by the ERG

The *de novo* analysis presented by the manufacturer uses a three health state Markov model (stable disease, progressive disease, and death). The stable disease state is subpartitioned into "On second-line treatment" and "Discontinued second-line treatment". Transition probabilities from stable disease to progressive disease and death were informed by independent parametric survival curves for PFS and OS fitted to patient level data from each separate arm of the VELOUR trial. Transition

probabilities within the stable disease state were estimated from independent parametric functions fitted to "Time to second-line treatment discontinuation" and "Time to early second-line treatment discontinuation" patient level data from the VELOUR trial.

The model considers aflibercept + FOLFIRI versus FOLFIRI alone as second-line treatment of mCRC 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. In addition to the general population of patients previously treated with an oxaliplatin-based regimen for mCRC who are eligible for a second-line treatment, the manufacturer also reported results for two subgroups: i) a *pre-specified* subgroup of patients with metastases limited to the liver; and ii) a *post-hoc* subgroup of patients that excluded those patients who had received oxaliplatin-based adjuvant therapy and relapsed on or within 6 months of this, i.e. those who had received adjuvant chemotherapy only.

Resource use data was collected in an UK observational study for secondary care and a KOL survey for adverse events treatment, and community and social care resource use. Total costs were calculated by applying unit costs from national published sources to the resource use estimates, and used to inform the economic model. Utility estimates for each health state were obtained from an observational study of patients with mCRC undergoing second-line treatment in the UK and Netherlands, the mCRC Utilities study.

Table 5.1 provides a summary of the manufacturer's economic evaluation, with justifications for key aspects and signposts to the relevant sections of the MS. The ERG has considered the methods applied in the manufacturer's economic evaluation in the context of a detailed checklist[36], reported in Appendix 2.

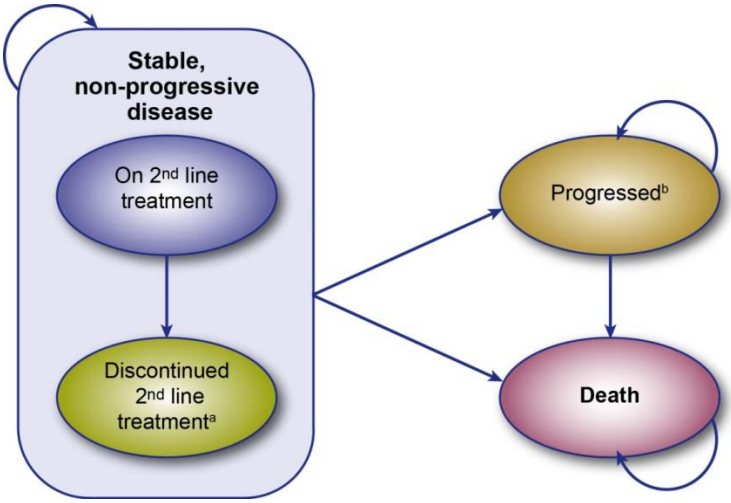
5.2.1 Model structure

The *de novo* analysis presented by the manufacturer uses a three health state Markov model (Figure 5.). The three states are: i) Stable (non-progressive) disease; ii) Progressive disease (PD); and iii) Death. The stable state is further subpartitioned into "On second-line treatment" and "Discontinued second-line treatment". Patients enter the model in the stable state and the second-line treatment substate. At each 14 days cycle, patients can either remain on second-line treatment or can discontinue second-line treatment while continuing to remain stable. Patients can also transition through the

model to PD or Death from either of the stable states. No reversion from the PD state to the stable state is possible.

Further lines of treatment are permitted after discontinuation of second-line treatment (both in the discontinued second-line treatment substate and PD state). The probability of receiving subsequent therapy and the type of subsequent therapy was assumed to be the same for each treatment cohort.

Figure 5.1 Model structure (Figure B-10, of the MS)



[a] Patients who stopped second-line treatment due to adverse events or physician or patient preference.

[b] Progressive disease - Implies at least a 20% increase in the sum of the longest diameter of target lesions or unequivocal increase in the size of non-target lesions or the appearance of 1 or more new lesions. Includes both third-line and best supportive care patients who can be differentiated by costs and utilities.

Life years were generated by the model according to the proportion of patients in stable and PD states at each cycle and summing these over the modelled time horizon. Quality of life was quantified by applying utility weights to each model state in order to estimate quality-adjusted life years. Utility decrements were applied to patients on treatment to reflect the impact of adverse events from second-line treatment on quality of life. Costs were assessed from an NHS and personal and social services (PSS) perspective and incorporated acquisition, administration and management costs of the alternative regimens, as well as subsequent lines of treatment, and adverse events.

Table 5.1 Summary of the manufacturer's economic evaluation, with justifications for key aspects and signposts to the relevant sections of the MS

	Approach	Source / Justification	Location in MS
Model	Cost-effectiveness (cost-utility) analysis using a Markov model.		Section 7.2.2; P80
States and events	The model contains 3 states: stable disease, progressive disease, and death. The stable disease state is subpartitioned into "On second-line treatment" and "Discontinued second-line treatment".	Health states are aligned with one of the key treatment objectives (prevent or delay disease progression) and with the primary and secondary endpoints of the VELOUR trial, OS and PFS respectively. The model structure is common in economic evaluation of oncology treatments, and is also consistent with those used in previous health technology assessments for the treatment of mCRC. The stable disease state was split, so that differences in costs and utilities during and after treatment whilst in PFS could be incorporated in the analysis.	Section 7.2.2; P80 Section 7.2.3; P81
Comparators	Aflibercept plus FOLFIRI was compared to FOLFIRI alone.	FOLFIRI or irinotecan monotherapy is recommended as second-line therapy for advanced or metastatic mCRC following first-line oxaliplatin-based therapies. Irinotecan monotherapy was not considered a relevant comparator, since it is not as widely used as FOLFIRI. Furthermore, irinotecan monotherapy is usually reserved for patients who cannot tolerate further fluorouracil-based therapy. Aflibercept is only licensed to be used in combination with FOLFIRI, which implies that patients eligible for aflibercept should be suitable for second-line FU-based combination chemotherapy. Since aflibercept's license requires first-line treatment with oxaliplatin-based regimens, second-line oxaliplatin-based regimens are not considered relevant comparators. Bevacizumab and cetuximab in combination with FOLFIRI were excluded from the comparison, as they are not currently recommended by NICE for the second-line treatment of mCRC. Panitumumab in combination with FOLFIRI was excluded as a	Section 2.7; P29

	Approach	**Source / Justification**	**Location in MS**
		treatment for only KRAS wild type patients, therefore a subset of the patients licensed for treatment with aflibercept. Moreover, panitumumab is not currently recommended by NICE for the treatment of mCRC after first-line chemotherapy.	
Subgroups	The following subgroups are considered in the MS: – Patients with liver metastases only (n=153 on aflibercept + FOLFIRI; n=146 on FOLFIRI alone); – *Post-hoc* subgroup that included the population that had not relapsed during or within 6 months of completion of adjuvant oxaliplatin-based therapy (n=552 on aflibercept + FOLFIRI and n=550 on FOLFIRI alone).	Patients with liver metastases only were a pre-specified group, and have been identified in previous NICE guidance as a clinically important subgroup in mCRC. Patients with rapid relapse after adjuvant therapy may have tumours with genetic changes that make them less sensitive to chemotherapy, and may not benefit from VEGF/chemotherapy. The treatment effectiveness was estimated similarly to the overall population, but using survival data (OS and PFS) from each subgroup only.	Section 3.5; P50 Section 6.3.7; P50 Section 7.9.1; P148-149 Section 7.9.2; P149 Section 7.9.3; P149 Section 7.9.3; P149
Natural History	Based on a Markov model. Transition probabilities between states were based on parametric survival functions fitted to patient-level data from the VELOUR trial.	Patients entered the model at the stable disease on second-line treatment state. Transition probabilities for disease progression and death were estimated from parametric functions fitted to PFS and OS patient level data from the VELOUR trial. Transitions probabilities within the stable disease state were estimated from parametric functions fitted to “Time to second-line treatment discontinuation” and “Time to early second-line treatment discontinuation” patient level data from the VELOUR trial.	Section 7.2.2; P80 Section 7.2.3; P81 Section 7.2.5; P82
Treatment effectiveness	Clinical outcomes included PFS and OS. OS was extrapolated by fitting independent parametric survival curves to the Kaplan Meier (KM) estimates for each treatment arm from the VELOUR trial. PFS was estimated by fitting independent parametric survival	All efficacy data was taken from the VELOUR trial. Extrapolation of OS was required as not every trial participant had experienced the event (death) at the end of follow-up, and therefore OS could be underestimated if only observed data had been included in the analysis. Nevertheless, OS (KM) curves are stated to be quite complete.	Section 6.5.3; P53-63 Section 7.3.1; P84-89

	Approach	**Source / Justification**	**Location in MS**
	curves to the KM estimates for each treatment arm from the VELOUR trial.	The use of parametric survival functions to estimate PFS was justified based on the existence of clear steps on the KM curves. The parametric survival curves smoothed those steps.	
HRQoL	Utilities for each health state in the model were estimated from EQ-5D administered in a separate observational study (the mCRC Utilities study), and assigned to each health state in the model. Utility estimates for main health states were assumed to be the same regardless of treatment, although separate treatment specific adjustments were applied to reflect the different adverse event rates.	EQ-5D was administered in the mCRC Utilities study ([redacted]), an observational non-interventional cross-sectional study in patients with mCRC had received an oxaliplatin-based therapy in first-line of treatment, and were receiving subsequent lines of therapy (including BSC). Utility weights were estimated by application of time trade-off tariffs[37] . The mCRC Utilities study population was assumed to be reflective of patients in the UK routine clinical practice. Utility data in the study was collected for the following health states: – Stable currently receiving second-line therapy ([redacted]) – Stable currently receiving third-line therapy or BSC following discontinuation from second-line therapy due to an adverse event ([redacted]). – Progressed on or following second-line therapy and currently receiving third-line or subsequent lines of therapy or BSC ([redacted]) A review of previously published utility values was undertaken by the manufacturer, but not used to inform the model, as the mCRC Utilities study was considered by the manufacture a more appropriate source of HRQoL evidence.	Section 7.4.6; P99-111 Section 7.4.7; P111 Section 7.4.9; P112-115
Adverse events	Adverse events were included if they were grade 3/4 and there was incidence greater than 5% in any of the treatment arms in VELOUR. Additionally, six rarer adverse events that were considered relevant by the clinical advisory board were also included.	In the VELOUR trial, the incidence of adverse events, including grade 3-4 adverse events, was increased in the aflibercept arm. Adverse events led to permanent discontinuation from study treatment in 26.8% of patients in the aflibercept arm and 12.1% of patients in the control arm. The impact of AE's was reflected on : – HRQoL	Section 6.9.2; P68-71 Section 6.9.3; P72 Section 7.3.1; P89-90 Section 7.4.9; P114-115

	Approach	**Source / Justification**	**Location in MS**
		Overall disutility experienced by a patient on a given second-line treatment was multiplied by the cycle probability of experience any of the included AEs, applied to patients on stable, on second-line treatment, state. – Costs Weighted average of costs of AEs treatment per treatment arm in the VELOUR trial was multiplied by the cycle probability of experience any of the included AEs, applied to patients on stable, on second-line treatment, state.	Section 7.5.7; P132-133
Resource use and costs	Cost categories included, drug acquisition, administration and medical management for treatments under comparison, as well as subsequent lines of therapy; and treatment of adverse events.	Drug acquisition costs (excluding Aflibercept) were taken from the British National Formulary[38]. The cost of Aflibercept is based on a patient access scheme proposed by the manufacturer which has been agreed by the Department of Health. Resource use items for the main health states were obtained from a retrospective observational study in mCRC, the UK mCRC Resource Use study, and a clinical oncology expert survey, the KOL survey. Both sources were UK specific. Resource use was costed using national unit costs from published sources.	Section 7.5.5; P129-131 Section 7.5.6; P131-132 Section 7.5.7; P132-133 Section 7.5.8; P133-134
Discount rates	Costs and benefits were discounted at 3.5% per annum.	In accordance with the NICE reference case.	Section 7.2.6; P82
Sensitivity analysis	Sensitivity analysis included univariate deterministic sensitivity analysis performed on a series of model parameters, and PSA. A series of scenario analyses was also performed to test structural assumptions.	In accordance with the NICE reference case, with the exception of PSA. Despite PSA having been performed, the actual cost-effectiveness results reported in the MS are based on the deterministic analyses. The reported outputs of the PSA were the probabilities of aflibercept + FOLFIRI being cost-effective at a threshold of £50,000 per QALY and the respective cost-effectiveness acceptability curves) for both treatment strategies. Representations of the PSA simulation results on the cost-effectiveness plane were also presented.	Section 7.6.1-3; P135 Section 7.7.7; P145 Section 7.7.8; P145-146 Section 7.7.9; P146-147 Section 7.7.10; P147-148 Appendix16; P59; P61-62
BSC, Best supportive care; EQ-5D, EuroQol Group QoL questionnaire; HRQoL, Health related quality of life; KM, Kaplan-Meier; NHS, National Health Service; PD, progressive disease; PAS, patient access scheme; PSA, probabilistic sensitivity analysis; VEGF, vascular endothelial growth factor.			

5.2.2 The manufacturer's economic evaluation compared to the NICE reference case checklist

The manufacturer's base case economic evaluation meets the criteria of the NICE reference case checklist (Table 5.2). No evidence synthesis was undertaken of efficacy outcomes as only one study of aflibercept in the second-line treatment of mCRC was identified in the systematic review. The ERG has been informed by our clinical advisor that other treatments (e.g. panitumumab, cetuximab and bevacizumab) are being used in clinical practice but it is not clear that this use could be considered representative of routine clinical practice, nor are any of these other treatments approved by NICE for this indication.

Table 5.2 A comparison to the NICE reference case

Attribute	Reference Case	Included in MS	Comment on whether *de novo* evaluation meets requirements of NICE reference case
Comparator(s)	Alternative therapies in the NHS, including those currently regarded as current best practice	Yes	The comparator in the model is FOLFIRI alone. It is unclear whether panitumumab, cetuximab and bevacizumab are used routinely in clinical practice.
Perspective - costs	NHS and PSS	Yes	
Perspective - benefits	All health effects on individuals	Yes	
Time horizon	Sufficient to capture differences in costs and outcomes	Yes	The economic model follows a time horizon of 15 years. Less than 0.01% of patients are expected to survive beyond this period, using the manufacturer's chosen extrapolation methods.
Synthesis of evidence on outcomes	Systematic review	NA	Only one study of aflibercept in second-line treatment of mCRC.
Outcome measure	QALYs	Yes	QALYs were calculated using utilities from a study assessing EQ-5D collected from mCRC patients. Utilities were applied to health states. The time in each health state was calculated using a Markov model with health state transitions calculated from the VELOUR trial.
Health states for QALY measurement	Described using a standardised and validated instrument	Yes	Derived from EQ-5D data.
Benefit valuation	Time Trade Off or Standard Gamble	Yes	Time Trade Off
Source of preference data	Representative sample of the public	Yes	Societal tariffs from EQ-5D.
Discount rate	3.5% on costs and health benefits	Yes	Costs and benefits have been discounted at 3.5% per annum.
Equity weighting	No special weighting	Yes	No special weighting undertaken.
Sensitivity analysis	Probabilistic sensitivity analysis	Yes	Probabilistic sensitivity analysis was undertaken.

EQ-5D, EuroQol Group QoL questionnaire; MS, Manufacturer's submission; NHS, National Health System; PSS, Personal social services; QALY; Quality adjusted life year.

5.2.3 Population

The economic evaluation is based on the VELOUR trial and thus the model population reflects the same characteristics as the trial population. As discussed in section 4.2, the demographic and disease

Superseded – see erratum

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 the UK between 2007 and 2009, an average 72% of colorectal cancer cases were diagnosed in people aged 65 years and over, whilst in the trial only [redacted] of the aflibercept group and 38.9% of the placebo group were aged over 65 years; the proportion of patients in the trial with an ECOG PS of 2 (2.2%) was, according to our clinical advisor, lower than seen in standard UK practice; and the proportion of patients who had only one primary site was higher in the trial population than in patients seen in standard UK practice. Furthermore, in the UK the median age of mCRC patients at diagnosis has been estimated to be 70 years[10], whereas the median age at baseline in the VELOUR trial was 61.0 for both treatment arms. 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.

To test the effect of treating a population more representative of mCRC patients treated in UK clinical practice, further exploratory analyses were undertaken by the ERG to evaluate the effect of treating an older population with a lower health-related quality of life in Section 6.2.

The model allows for the evaluation of the overall population and two subgroups, the pre-specified liver metastases only population and a population of patients who progressed more than 6 months after receiving adjuvant oxaliplatin therapy (i.e. excluding patients who had had adjuvant chemotherapy). Data specific to the overall population and each of the subgroups were used in the economic evaluation, this included information about OS, PFS, treatment duration, the number of treatment administrations per cycle dose intensity per administration, weight and body surface area (BSA) distributions, and AE probabilities. Using the manufacturer's base case assumptions the ICER in the overall population was £36,294 per QALY. For the subgroups ICERs were £30,474 per QALY for the liver metastases only population and £32,480 per QALY for the excluding adjuvant chemotherapy population.

5.2.4 Interventions and comparators

5.2.4.1 Appropriateness of comparators

The only relevant comparator was identified by the manufacturer as FOLFIRI alone. The ERG have previously noted that it is likely that other treatments are currently used in UK clinical practice, including, panitumumab, cetuximab and bevacizumab. It is unclear the extent to which these other

treatments are being used in this indication and hence cannot be considered by the ERG to represent routine clinical practice. Panitumumab, cetuximab and bevacizumab have also been previously evaluated by NICE for use as a second-line treatment in mCRC and have been rejected. Consequently, the ERG considered FOLFIRI alone to be the most appropriate comparator for the specific decision problem under consideration within this appraisal, i.e. the second-line treatment of mCRC for patients who have previously received oxaliplatin-based chemotherapy (see section 3.3).

5.2.5 Perspective, time horizon and discounting

The economic evaluation was conducted from an NHS and personal and social services perspective. The model had a time horizon of 15 years, which was stated to represent a lifetime horizon. Inevitably this required the extrapolation of the observed data from the RCT over a significantly longer period. The median OS for patients on second-line treatment for mCRC ranges from 10.5 to 18 months in a number of clinical trials[2-4] and one observational study[5]. Only 0.01% patients in the model were predicted to remain alive beyond 15 years, thus this would usually be considered an appropriate time horizon. Nevertheless, the ERG is concerned that mean OS gains may have been overestimated through the assumptions applied in the economic model related to the extrapolation (see section 5.2.6.3).

A 3.5% per annum discount rate was applied for costs and outcomes, in line with NICE guidance.

5.2.6 Treatment effectiveness and extrapolation

In the manufacturer's model it is assumed that treatment affects both PFS and OS independently by changing the transition probabilities between the health states: stable, PD and death. The transition probabilities were estimated by fitting independent parametric survival curves to empirical data from each arm of the VELOUR trial. The use of parametric curves enabled the observed OS data to be extrapolated beyond the follow-up period of the trial data in order to estimate mean survival times required for the cost-effectiveness analysis. The use of parametric survival functions to estimate PFS, despite the limited censoring, was justified by the manufacturer based on the need to smooth the steps on the Kaplan Meier (KM) curves. It is not clear to the ERG why the manufacturer deemed it necessary to smooth these steps. However, since the KM functions were either complete (or nearly complete), the use of parametric functions rather than the KM data itself is unlikely have had an important effect on the ICER results.

Superseded – see erratum

5.2.6.1 PFS estimation

The health state transitions from stable to PD were calculated from the PFS data of the VELOUR trial. The PFS results have been previously reviewed in section 4.2.3. The manufacturer's base case economic analysis uses the IRC assessment of lesions rather than the investigator's assessment of lesions. The investigator's assessment of lesions was tested by the manufacturer in a scenario analysis using Weibull curves the ICER was £36,410 per QALY. The transition probabilities to PD were calculated from independent Weibull curves for both the aflibercept + FOLFIRI and FOLFIRI arms based on statistical goodness of fit statistics (e.g. using AIC and BIC tests for the aflibercept arm (Table B20, MS) and justification from the NICE DSU that "in the absence of strong justification it is not appropriate to apply different functions to two treatment groups for the same endpoint"[39] The manufacturer undertook several sensitivity analyses on the extrapolation of PFS. The ICER ranged from £34,935 to £36,410 per QALY and was considered by the ERG to be appropriate and robust.

5.2.6.2 OS extrapolation

The base case model uses transition probabilities estimated from independent log-logistic curves to extrapolate the OS data from the trial for each treatment group. As described in Section 4.2.3 the selection of the log-logistic function was justified based on the AIC and BIC tests. Proportional hazards models were rejected by the manufacturer in their base case analysis and scenario analyses, as the proportional hazard assumption was considered not to hold by the manufacturer. Nevertheless, the economic model allows using the proportional hazards model as an alternative extrapolation method. Although these were not reported within the manufacturer’s report, when the ERG used this alternative extrapolation model, the ICER increased to £58,784 per QALY, with the difference being driven mainly by a 43.8% reduction in QALY gain compared to the manufacturer’s base case.

Importantly, regardless of which OS extrapolation method is selected, independent parametric curves or a proportional hazards model, the underpinning assumption of either approach is that there is a continued treatment effect on OS beyond the period for which there is empirical data on OS from the trial. The main difference between the models is that in the proportional hazards model the HR between treatments is constant, whereas for the independent curves model, the HR varies over time. 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

Superseded – see erratum

gains (i.e. based on the observed data and the projected survival estimates including the extrapolation period) with aflibercept + FOLFIRI.

The assumption that the OS curves continue to diverge 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).

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[33]. Given this uncertainty, it is important that this is adequately reflected in the assumptions used to inform the extrapolation. The NICE methods guide for Technology Appraisal states that where uncertainty exists related to the extrapolation assumptions, alternative scenarios should be explored using different assumptions[40] (i.e. assuming that treatment effects continue in a similar manner to the observed period; assuming that any benefits immediately cease at the end of the observed period or assuming that any benefits decrease over the extrapolation period. However, by fitting independent survival curves in the base case analysis, the assumed treatment effects being applied over the extrapolation period remain largely implicit in the manufacturer's submission. A further consequence of this is that it becomes more challenging to investigate these alternative scenarios which can only be considered by altering the survival functions directly (e.g. assuming the respective survival curves continue to diverge, remain parallel or begin to converge).

Importantly, despite the high levels of uncertainty in the OS data, the manufacturer did not consider the full range of alternative assumptions and did not consider the possibility that the OS may converge over the extrapolation period (i.e. based on the shape of the respective PFS curves). These alternative scenarios are explored by the ERG in the additional analyses reported in Section 6.

5.2.6.3 Treatment duration

Treatment with aflibercept + FOLFIRI and FOLFIRI in the VELOUR trial was repeated every 2 weeks until progression, unacceptable toxicity or choice. Treatment duration was modelled based on "time to treatment discontinuation" (TTD) time to event data from the VELOUR trial. Parametric survival functions were fitted to "time to second-line treatment discontinuation" and "time to early second-line treatment discontinuation" (prior to disease progression), and used to determine the time spent in the two stable, non-progressive disease health substates. The distinction between the two

TTD curves was explained by the manufacturer in response to clarification questions (question A20). "Time to second-line early discontinuation" referred to patients who had discontinued early due to adverse events or patient choice, and had not experienced progression, whereas "time to second-line treatment discontinuation" referred to treatment discontinuation due to any cause, including progression. The proportion of patients in the trial who had discontinued before progression was [redacted] and [redacted] of all patients in the ITT patient population for the aflibercept + FOLFIRI and the FOLFIRI arms respectively. The same proportion of patients in the model was assumed to discontinue early, i.e. follow the "Time to second-line early discontinuation" curve.

Independent parametric survival curves were fitted to empirical data on treatment discontinuation for each treatment group data, and for both sets of TTD curves, based on a similar rational as for OS and PFS. The survival functions were fitted in the same way as for the PFS and OS endpoints (described above). In the MS, the AIC and BIC estimates are reported in tables B21 and B22 (MS), and visual inspection plots are presented in figure B-13 and B-14 (MS).

The survival distributions for discontinuation selected for the base case scenario were Weibull functions. The manufacturer's interpretation of the curves statistical fit appeared to be appropriate, but the ERG cannot fully comment on the appropriateness of the visual fit assessment, as the graphical plots only included the representation of the Kaplan-Meier curves and the fitted Weibull functions. The ERG believes this is unlikely to have led to an alternative selection of parametric functions, given the completeness of the treatment discontinuation data. Furthermore, the ERG found the model to be robust to alternative assumptions regarding the selection of survival functions to model treatment discontinuation. However, it should be noted that the economic model does not allow the use of Gompertz or Exponential fitted curves to model treatment discontinuation.

It should be noted that the probability of being in the stable, on second-line of treatment, health state in the model did not affect the acquisition and administration cost of second-line treatment, as these costs were not assigned to any model state. In the manufacturer's base case the acquisition and administration costs of second-line treatment were applied to every patient in the model for a fixed number of cycles (correspondent to the mean overall duration of the combination treatment in each treatment arm). The ERG is concerned that this may have underestimated costs in both treatment groups. This issue is further discussed in Section 5.2.8.2, and formally addressed in Section 6.

5.2.7 Health related quality of life

HRQoL was reflected by assigning utility estimates to each of the model health states and substates. Utilities for each health state in the model were estimated from EQ-5D collected (based on cross-sectional administration of the questionnaire) in the mCRC Utilities study, and assigned to each health state in the model.

The manufacturer's mCRC Utilities study was an observational non-interventional cross-sectional study in patients with mCRC ([redacted]) that had received an oxaliplatin-based therapy in first-line of treatment, and were receiving subsequent lines of therapy (including BSC). Only patients who had an ECOG PS equal or lower than 2 at the time of mCRC diagnosis were enrolled. The study was conducted in the UK ([redacted]) and Netherlands ([redacted]), and was assumed by the manufacturer to be reflective of patients in UK routine clinical practice.

Patients in the study were assigned to the following health states, according to data on their medical records:

- Stable currently receiving second-line therapy ([redacted])
- Stable currently receiving third-line therapy or BSC following discontinuation from second-line therapy due to an adverse event ([redacted]).
- Progressed on or following second-line therapy and currently receiving third-line or subsequent lines of therapy or BSC ([redacted])

The results obtained in the study, and utility estimates applied to the economic model are shown in Table 5.3. In addition, a utility of zero was assigned to the death state. The utility estimates in each state were applied to the proportion of patients within each health state during each cycle of the model, in order to determine the overall QALY estimates for each treatment strategy.

Table 5.3 Summary of utility weights in the economic evaluation (model inputs emboldened)

State in the mCRC Utilities study	Study Mean utility value* (SD)	State in the model	Model Mean utility value (SE)
Stable currently receiving second-line therapy	[redacted]	Stable, on second-line treatment	[redacted]
Stable currently receiving third-line therapy or BSC following discontinuation from second-line therapy due to an adverse event	[redacted]	Stable, discontinued second-line treatment	[redacted]
Progressed on or following second-line therapy and currently receiving third-line or subsequent lines of therapy or BSC	[redacted]	Progressive disease	[redacted]
* Results refer to an interim analysis performed in the 21st December 2012 database. SD, standard deviation; SE, standard error.			

The utility estimate obtained for "Stable currently receiving third-line therapy or BSC following discontinuation from second-line therapy due to an adverse event" was not used in the model due to the small number of patients in this health state ([redacted]). The manufacturer assumed the same utility weight for both the stable disease substates (on second-line treatment and discontinued on second-line treatment"). Health state utilities were assumed to be independent from the type of second-line treatment (aflibercept + FOLFIRI or FOLFIRI alone). AEs rates and weighted average disutilities specific to each treatment were applied, for patients on the stable, on second-line treatment state.

The ERG has some concerns regarding the generalisability of the mCRC Utilities study to the mCRC patient population treated in UK routine clinical practice. Even though the study included a majority of UK mCRC patients ([redacted]), the patient median age was [redacted] years and the proportion of patients with an ECOG PS equal to 2 was only [redacted], and is not considered to be reflective of the patients in actual UK clinical practice. The regression analyses performed in the mCRC Utilities study did not allow any conclusion to be drawn on the potential predictive power of ECOG PS and age.

Furthermore, the study was relatively small, and even produced counter-intuitive results in one instance. In the study's UK only subpopulation the mean utility estimate for the patients in progression was greater than for the patients considered to be stable, on second-line treatment.

5.2.7.1 Other sources of HRQoL data

The manufacturer identified one ongoing clinical trial study (the ASQoP trial, described in Section 4.1.7) that is currently collecting HRQoL data for patients receiving aflibercept + FOLFIRI using a variety of instruments (EQ-5D, EORTC QLQ-CR30, and EORTC QLQ-CR29), and aims to enrol 900 patients. The only results presented by the manufacturer for this study refer to the mean baseline utility weight of [redacted] patients ([redacted], SD = [redacted]). According to the MS, more data are expected to become available within the course of the appraisal process.

A systematic review was also conducted by the manufacturer to identify utility estimates potentially relevant to the economic evaluation. The systematic review identified six utility studies, one study reporting SF-36 data, and documentation (including ERG reports and NICE guidance) relative to five NICE HTAs that reported utility estimates in mCRC. The studies and HTA documentation were assessed by the manufacturer for relevance to this appraisal and consistency with the NICE Reference Case. The manufacturer did not include any of these utility estimates in the model.

The utility estimates for stable disease from the mCRC Utilities study were considered by the manufacturer to be within the range of estimates reported in relevant published literature (0.73-0.81).[10, 41, 42] No comparison between study results and other values found in the external literature for PD was discussed. The utility estimates considered appropriate by the manufacturer for the PD state identified in the literature ranged between 0.68-0.69. [10, 43, 44] The manufacturer considered progressed on or following second-line therapy utility estimates from the mCRC Utilities study to be reflective of HRQoL in the PD state for patients who have progressed and whose symptoms are still relatively well controlled either through further lines of active treatment or BSC. Despite this, the manufacturer recognised that utilities are likely to decline in PD state, especially near the end of life.

The study conducted by Best *et al* 2010[45] was considered inappropriate by the manufacturer, as data referred to a mixture of patients receiving adjuvant chemotherapy, in remission, and with metastatic disease. However, utility estimates were reported separately for metastatic disease (stable and progressive). The manufacturer also questioned the coherence of the study, since patients value their own health state higher than they valued all hypothesised health states characteristic of the same disease. This notwithstanding, the ERG would like to highlight that utility estimates from the study have been used in previous evidence review reports in mCRC.[46] Despite the study utility estimates being much lower than those applied in the economic model, the manufacturer did not explicitly compare them to the mCRC Utilities study or suggest any rationale for the difference. A selection of relevant results from the Best *et al*., 2010,[45] study are shown in Table 5.4.

Table 5.4 Selection of utility estimates from Best *et al.*, 2010 [45]

	TTO scores			
Health state	**Patients (n=49)**		**Community (n=49)**	
	Raw	**Adjusted***	**Raw**	**Adjusted***
Remission	0.87	0.83	0.83	0.82
Metastatic, stable	0.46	0.40	0.54	**0.51**
Metastatic, progressive	0.38	0.37	0.21	**0.21**
Current health	0.82	—	0.94	—

TTO, time trade-off.

Utility estimates used in CG 131[46] are emboldened.

* Adjusted TTO values based on regression analysis including dummy variables for seven chronic health states and covariates age, education and current health. Adjusted means are means are based on 60 year-old patient with mean education and mean TTO for current health.

In another HTA, namely the appraisal of bevacizumab in combination with oxaliplatin and either fluorouracil plus folinic acid or capecitabine for the treatment of mCRC (TA212),[42, 43] the manufacturer considered the utility estimates for first-line chemotherapy, and disease progression, to be appropriate for the stable non-progressive disease state, and progressive disease states respectively. However, it should be clarified that the utility estimates from this appraisal mentioned in the aflibercept MS refer to estimates reported by Roche in the bevacizumab MS. The ERG who reviewed the bevacizumab MS considered these estimates to be poorly referenced, and were unable to verify them. Furthermore, the ERG for the bevacizumab technology appraisal expressed concerns relating to the clinical plausibility of the post-progression utility estimate applied in Roche's economic model,[43] as utility is likely to decline in the last few weeks of life. Uncertainty at this level was explored by the ERG for the bevacizumab technology appraisal by simultaneously varying utilities values 20% above and below the point estimate of 0.68 (though the ERG highlighted that utility estimates were unlikely to increase). The utility estimates in the bevacizumab appraisal and identified by Sanofi as relevant are shown in Table 5.5.

Table 5.5 Summary of utility estimates in TA212[43]

Health state	**Utility estimate**
First-line chemotherapy	0.77
Second-line PFS	0.73
With disease progression	0.68*

* Value applied in the model. The value in the original MS was 0.67.

Superseded – see erratum

The ERG is also concerned that utility estimates in the mCRC Utilities study (and most utility studies identified) appear to be generally high, when compared to other real world estimates of utility in colorectal cancer, and even with the UK general population estimates. In a UK based study of EQ-5D index scores by health conditions, the utility weight estimated for patients with colon cancer (not necessarily metastatic) was 0.673 (mean age of the patients from which this estimate was obtained 64.7 years old), or using a different health condition classification system, 0.676 (Malignant neoplasm of the colon, patient mean age equal to 42.8 years old).[47] The UK general population utility has been estimated as 0.8 for the 55-64 year old age bracket, declining to 0.78 in the next age bracket (65-74 year old) and to 0.65 for age greater than 75 years old.[48] 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,[48] in the Health Survey for England.[49]

The manufacturer recognised that utility estimates discussed in the MS are relatively high when considering the age and health status of the patients. The ERG agreed, particularly in the case of HRQoL in progressive disease.

5.2.7.2 Post-progression HRQoL

The assumptions surrounding the utility weight attached to the PD health state are a source of concern to the ERG, especially as this parameter is a key driver of cost-effectiveness in the economic model. The post-progression utility used in the model is relatively high ([redacted]) compared to utility estimates in the external literature (range: 0.21-0.69)[10, 42-45]. In a previous HTA report[1] it has been commented that even a utility weight of 0.69 for patients in PD states is likely to be an overestimate, as the time from measurement of HRQoL in progression to death was unknown, and so the estimate may not have been reflective of the full period in progression. The utility estimate obtained in the mCRC Utilities study was used in the model to be reflective of utility in PD state throughout time. 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. In order to assess the validity of the assumption the ERG asked the manufacturer for more detail regarding time since progression for patients identified as on progression on second-line treatment in the mCRC Utilities study. The mean and median time from progression in second-line of treatment to informed consent was [redacted] and [redacted] respectively. It is unclear why mean and median time from progression in second-line does not include [redacted] of the patients. Moreover, this study may not

have captured patients with more severe disease due to participation bias. Given the time horizon in the model, HRQoL deterioration associated with ageing of the cohort is also likely to occur, further reducing the utility in PD.

5.2.7.3 HRQoL adjustment for adverse events

For the purpose of the economic evaluation, the manufacturer included Grade 3 and 4 adverse events that affected more than 5% of the population in the VELOUR trial, together with six additional rarer adverse events that were considered relevant by a clinical advisory board. The impact of adverse events on HRQoL per patient was obtained by multiplying the overall disutility experienced by a patient on a given second-line treatment by the probability of incurring any of the included AEs at each cycle, and applying this to all patients in the stable, on second-line treatment, state. Overall disutility was a weighted average of the disutilities presented in Table 5.6 by the distribution of the specific adverse events within each treatment arm and the mean duration of the adverse events (assumed to equal the length of treatment in years). The probability of experiencing an adverse event was estimated by converting the event rate for each treatment arm to a cycle probability.

Table 5.6 Data used to estimate disutility due to adverse events

Adverse event	Disutility	Distribution of AEs %		Mean length of treatment (years)	Number of events	
		Aflibercept + FOLFIRI	FOLFIRI		Aflibercept + FOLFIRI N=[redacted]	FOLFIRI N=[redacted]
Diarrhoea	–0.103	15.6	10.6	0.12	[redacted]	[redacted]
Hypertension	–0.069	16.5	2.1	0.16	[redacted]	[redacted]
Fatigue	–0.115	10.2	10.6	0.42	[redacted]	[redacted]
Stomatitis	–0.151	9.4	6.1	0.21	[redacted]	[redacted]
Abdominal pain	–0.069	2.8	2.9	0.15	[redacted]	[redacted]
Vomiting	–0.103	1.9	5.2	0.02	[redacted]	[redacted]
Dehydration	–0.103	3.0	1.7	0.03	[redacted]	[redacted]
Decreased appetite	–0.069	2.4	2.5	0.38	[redacted]	[redacted]

Adverse event	**Disutility**	**Distribution of AEs %**		**Mean length of treatment (years)**	**Number of events**	
Nausea	–0.048	1.1	4.2	0.04	[redacted]	[redacted]
Neuropathy peripheral*	–0.116	0.7	0.6	0.44	[redacted]	[redacted]
Urinary tract infections*,**	–0.195	0.5	1.0	0.11	[redacted]	[redacted]
Hand-foot syndrome*	–0.116	1.8	0.6	0.74	[redacted]	[redacted]
Gastrointestinal perforation*	–0.195	0.3	0.4	1.10	[redacted]	[redacted]
Haemorrhage*	–0.195	2.2	2.1	0.05	[redacted]	[redacted]
Febrile neutropaenia*	–0.15	3.0	2.1	0.02	[redacted]	[redacted]
Neutropaenia	–0.09	25.2	43.2	0.03	[redacted]	[redacted]
Asthenia	–0.115	3.4	4.0	0.44	[redacted]	[redacted]
Overall disutility					**Total patient-time on treatment*** **	
					[redacted]	[redacted]

* rare adverse events included in the model under advisement of clinical experts.

** From clarifications response, but initially referred to in the MS as infections.

***number of two-week model cycles

AEs, adverse events.

The disutility estimates for the AEs were based on the published literature, supplemented by clinical expert opinion, and were not CRC specific. Most disutility estimates selected by the manufacturer seem to be appropriate. The exceptions to this are three estimates, namely those referring to urinary tract infections, gastrointestinal perforation and haemorrhage, for which the appropriateness of the manufacturer's selected disutility value is debatable. First, the disutility associated with urinary tract infections in the MS does not correspond to the estimate in the referenced study (-0.073 rather than -0.195).[50] Secondly, the manufacturer could not find disutility estimates for gastrointestinal perforation and haemorrhage, and so assumed that this should be the highest value in the range of other disutilities. It is unclear whether this would be truly reflective of the disutility associate with these AEs. Although there is uncertainty regarding the appropriateness of these values, it is important to note that additional sensitivity analyses undertaken by the ERG has demonstrated that the model is robust to changes in these three disutility estimates and hence does not consider that this aspect is central to the validity of the overall cost-effectiveness estimates being reported.

Superseded – see erratum

The difference in overall disutility between treatment arms is relatively small, which explains the lack of impact on the ICER. In fact the model results are fairly insensitive to simultaneous changes in all disutility parameters, with the ICER increasing by only 0.16% even when disutility estimates are doubled.

However, when applying AEs related disutility in the model, the manufacturer appeared to divide the utility by the number of cycles 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.

5.2.7.4 Summary of Health Related Quality of Life

There are a number of issues related to the utility estimates used in the model and the uncertainty surrounding them that are likely to be overly optimistic towards aflibercept. Alongside the extrapolation of OS data for aflibercept + FOLFIRI, the ERG considers the utility assumptions to be a key driver of the cost-effectiveness and an area of significant uncertainty. The HRQoL estimates in the model seem to be generally high, and this is particularly relevant in the case of utility in PD state. Since the model is predicting a greater gain in OS for aflibercept + FOLFIRI compared to FOLFIRI than for PFS, the post-progression utility impact on the cost-effectiveness of the intervention is further compounded if OS is being overestimated. As it can be seen in Table 5.7, [redacted] of absolute QALY increment in the model was accrued post-progression.

Table 5.7 Summary of QALY by health-state (Table B41, MS)

Health state	QALY: aflibercept + FOLFIRI	QALY: FOLFIRI	Increment	Absolute increment	% absolute increment
Stable, non-progressive disease	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Progressive disease	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Adverse events	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Total	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

5.2.8 Resources and costs

A wide range of resources and costs were detailed in MS. These included: drug acquisition, administration and medical management for treatments under comparison, as well as subsequent lines of therapy; and treatment of adverse events. Management costs included supportive medications, physician and nurse visits (hospital and community), chemotherapy regimen change, imaging, laboratory tests, hospitalisations, palliative care, and personal and social care.

The manufacturer conducted a systematic review to identify resource use and cost studies relevant to the appraisal, namely those that related to mCRC and/or AEs related to second-line treatments in patients with mCRC. The manufacturer states that the systematic review identified two studies and three HTAs (documentation including ERG reports and NICE guidance) relevant to the appraisal, but the summary table in the MS refers to three studies and two HTAs. The studies and HTAs were assessed by the manufacturer for applicability to the evaluation, but it was not clear what the criteria for this assessment were. The HTAs costs were considered applicable, as well as one micro-costing study for FOLFIRI.[51] The latter was used by the manufacturer to compare with costs in the FOLFIRI arm, whilst no resource use or costs from the HTAs documentation was stated to have been used in the submission. No justification was presented for this.

Resource use associated with the management of patients with mCRC in the hospital setting was estimated through a retrospective observational study conducted by the manufacturer, the UK mCRC Resource Use study. Community-based care and personal and social care resource use was estimated via a survey of six UK clinical oncology experts, the KOL survey. The resource use associated with the treatment of adverse events and chemotherapy regimen change was also estimated in the KOL survey. Resource use in the end-of-life period ([redacted]) was excluded from the UK mCRC Resource Use study and the KOL survey. This exclusion was justified by the manufacturer based on the assumption that end-of-life care costs would be equal between treatment cohorts, which the ERG found appropriate.

The UK mCRC Resource Use study was set in [redacted] and included patients who received an irinotecan-based chemotherapy regimen, following prior treatment with an oxaliplatin-based regimen. All data were sourced retrospectively from eligible patients' medical notes

and electronic hospital systems by the routine clinical care team. Data collected in the study was classified according to disease phases, which were aligned with the health states in the model:

- Stable , on second-line treatment (patients receiving FOLFIRI): [redacted];
- Stable, discontinued second-line treatment: [redacted] for active treatment and BSC, respectively;
- Progressive [redacted] for active treatment and best supportive care, respectively.

Resource use per 2-week period was summarised separately for patients receiving active therapy and those receiving BSC as further lines of treatment (number and percentage of patients using the resource, mean number used, and standard deviation). The distribution, dosage, and treatment duration for subsequent therapies were also estimated from the Sanofi retrospective observational study, the UK mCRC Resource Use study. In the model, the proportions of patients on further lines of treatment was informed by data from the UK mCRC Resource Use study, and are shown in Table 5.8.

Table 5.8 Distribution of patients across further lines of treatment by model health state

	Stable, off 2nd line treatment	Progressive disease
Active treatment	[redacted]	[redacted]
BSC	[redacted]	[redacted]
BSC, best supportive care.		

Patients in the UK mCRC Resource Use study had a median age of approximately [redacted] years. No data regarding the distribution of patients across ECOG PS, and location and dissemination of metastases was provided in the MS. Therefore, the ERG was unable to fully assess whether the study population was comparable to the patient population in UK routine clinical practice, even if the study population appears to be younger. It is also unclear whether age and overall fitness is likely to affect resource use in patients with mCRC.

Costs were obtained by applying UK national tariffs to resource use estimates from the UK mCRC Resource Use study and KOL survey. The main sources of unit costs used in the submission were: NHS Reference Costs,[52, 53] Unit Costs of Health and Social Care[54], and British National Formulary (BNF) 62.[38] Details on unit costs can be found on Table B32 and B33 in the MS, and Table 1 in appendix 15 of the MS.

The ERG identified a number of issues regarding the estimation of resource use and the selection of unit costs that may impact on the cost-effectiveness of aflibercept. These issues are discussed below.

5.2.8.1 Drug acquisition costs

The price of aflibercept was stated in the MS to be awaiting confirmation by the Department of Health (DH). The list prices for the two aflibercept doses are shown in Table 5.9. The manufacturer also stated that the acquisition cost included in the model was calculated based on PAS prices that were approved by the DH during the course of this appraisal, rather than on the list price. The PAS is described as a simple discount, but the size of the discount (or the exact price of the drug with the PAS) could not be found within the MS. The ERG requested further details on the nature of the proposed PAS on several occasions, but did not receive any details. Hence, the ERG cannot verify whether the PAS has been appropriately incorporated or not. The value applied in the model can be found in Table 5.9. Assuming the prices in this table appropriately reflect the PAS agreement then the ERG would concur that this has been appropriately incorporated.

Table 5.9 List and model price of aflibercept

	List price	Price in the model	Estimated discount
Aflibercept 100 mg	£ 295.65	[redacted]	[redacted]
Aflibercept 200 mg	£ 591.30	[redacted]	
All prices exclude value-added tax.			

Replacing the prices used in the model with the list prices described in the MS increases the ICER from £36,294 to [redacted]. All analyses presented by the manufacturer (with the exception of budget impact estimation) were based on the values used in the model. The ERG also uses these prices (which we assumed to correspond to the PAS prices) on all additional analyses. Therefore, judgments on the relative cost-effectiveness of the intervention made on the ERG report are only applicable if a PAS that reduces the list price of aflibercept by [redacted] is in place.

5.2.8.2 Duration of second-line treatment

Most costs in the model were attached to each health state. Lifetime costs were accumulated by taking into account the proportion of patients in each health state of the model for each cycle. Conversely, acquisition and administration costs of second-line treatments (aflibercept + FOLFIRI or FOLFIRI alone) were calculated outside the patient flow in the model, i.e. the proportion of patients and length of time for which these costs were accrued did not depend on patient distribution in the model at each cycle . These costs were instead based on the mean overall duration of the combination

treatment observed in the VELOUR trial for each treatment arm, and were included for a fixed number of cycles for all patients in the model. The estimation of second-line drug acquisition costs was described by the manufacturer as being based on:

- Mean dose administered for each second-line treatment drug in the VELOUR trial – according to the manufacturer this accounted for dose reductions (due to toxicity) and delays;
- Drug cost per administration – the base case assumed wastage of second-line drugs, i.e. the cost of drug per dose administered was calculated assuming that unused drug in opened vials was discarded . The weight and BSA distributions of the Western European population in the VELOUR trial were used to estimate the proportion of patients receiving each vial combination.
- Duration of treatment - The mean duration of treatment for aflibercept + FOLFIRI and FOLFIRI alone were taken from the VELOUR trial.

To estimate the drug cost per 2-week model cycle, the manufacturer adjusted the drug cost per administration multiplying it by the ratio between the total number of administrations (for each drug component individually) and the total duration of treatment for the regimen (in 2-week model cycles). The mean drug cost per model cycle was applied in the model for a fixed number of cycles (correspondent to the mean overall duration of combination treatment in each treatment arm) and for every patient in the model. Table 5.10 details the elements used to estimate acquisition cost per cycle and per treatment arm.

Table 5.10 Costs and doses of second-line treatment - adapted from the MS (Table B35)

Items	Aflibercept + FOLFIRI	FOLFIRI
Administered dose		
Aflibercept	[redacted]	-
Irinotecan	[redacted]	[redacted]
Leucovorin	[redacted]	[redacted]
5-FU	[redacted]	[redacted]
Source	VELOUR trial, WE	VELOUR trial, WE
Drug cost (per model cycle): base case (includes wastage)[a]	[redacted]	[redacted]
Mean number of administrations		
Aflibercept	[redacted]	
Irinotecan	[redacted]	[redacted]
Leucovorin	[redacted]	[redacted]
5-FU	[redacted]	[redacted]
Source	VELOUR trial	VELOUR trial
Mean duration of treatment (model cycles)[b]	[redacted]	[redacted]
Source	VELOUR trial	VELOUR trial
Total cost (base case)	[redacted]	[redacted]

FU, fluorouracil; FOLFIRI, fluorouracil/leucovorin/irinotecan; VELOUR, Aflibercept Versus Placebo in Combination With Irinotecan and 5-FU in the Colorectal Cancer After Failure of an Oxaliplatin Based Regimen; WE, Western European population of the VELOUR trial.

[a] Mean cost per 2-week model cycle. Calculations include dose adjustment for patient weight and BSA, dose reductions and dose delays (e.g., for toxicity).

[b] The duration of treatment was defined as the overall treatment duration of the combination, i.e., from the first day of the cycle, where any part of the combination was first administered, to the last day of the cycle where the last component was administered.

The administration cost of the drugs was also assumed to be incurred for the mean overall duration of the combination treatment in the VELOUR trial.

The assumption of fixed treatment duration was justified by the manufacturer as necessary to reflect dose reductions and delays observed in the trial. However, the ERG believes any dose reductions and delays would have been more accurately reflected in the cost estimates, if drug cost per administration (including acquisition and administration costs) had been applied directly to the proportion of patients in each health state. In this way the appropriate costs are applied to the health states in the same way that the utilities are applied. The ERG considered this to be an error in the model and in Section 6 tested the effect of making this correction and included it in the ERG's preferred analysis.

5.2.8.3 Drug administration costs

Another issue was identified by the ERG regarding the administration costs of aflibercept; the manufacturer assumed that the administration costs were independent of the treatment arm. The manufacturer justified the assumption by stating that, despite aflibercept being administered over one hour, it would be administered alongside FOLFIRI. Therefore, no additional staff or inpatient admissions (and associated facilities) would be required to administer aflibercept.

Nevertheless, the administration of aflibercept as part of the aflibercept + FOLFIRI treatment requires the preparation of an additional infusion, which implies an incremental cost compared to FOLFIRI alone. In the ERG report of a previous HTA,[1] the pharmacy preparation of cetuximab and bevacizumab has been estimated to be £15 per infusion. The inclusion of this cost in the administration cost of aflibercept + FOLFIRI increased the base case ICER from £36,294 to £37,035 per QALY.

Another issue may arise if aflibercept administration implies an additional hour of infusion time when compared to administering FOLFIRI alone, i.e. if aflibercept is administered after or before FOLFIRI, instead of alongside it. In this case, the additional cost of the extra hour of infusion time for aflibercept would have to be considered in the model. The manufacturer estimated administration costs for both treatment strategies based on the Healthcare Resource Groups (HRG) code SB14Z for chemotherapy delivery on a day case and regular day/night setting from NHS reference costs.[52] The ERG agrees that this code is appropriate to cost the administration of FOLFIRI alone, but found no single code that could be used as an adequate alternative to estimate the cost of an extra hour of infusion time. Alternatively, a cost of £45 can be estimated from the difference between delivering a more complex and a simple parenteral chemotherapy at first attendance (code SB13Z and SB14Z respectively). The ERG believes this to be a reasonable estimate of the cost associated with one additional hour of infusion time, and found that an increase of £45 in the administration of aflibercept

+ FOLFIRI increased the base case ICER to £38,517 per QALY. HRG codes and respective unit costs can be found in Table 5.11.

Table 5.11 Chemotherapy delivery unit costs (day case and regular day/night) [52]

HRG code	HRG Description	National Average Unit Cost
SB12Z	Deliver Simple Parenteral Chemotherapy at First Attendance	£203
SB13Z	Deliver more Complex Parenteral Chemotherapy at First Attendance	£248
SB14Z	Deliver Complex Chemotherapy, including Prolonged Infusional Treatment, at First Attendance	£ 334
HRG, Healthcare Resource Groups.		

As it has been demonstrated, the model is sensitive to assumptions on additional administration costs for aflibercept + FOLFIRI. It is likely that this impact is further increased, once alternative assumptions are considered for treatment duration. The ERG exploratory analyses in Section 6 will aim to better characterise uncertainty at this level.

5.2.8.4 Key Opinion Leader Survey

In the KOL survey, 6 clinical oncology experts were asked to complete a resource use questionnaire which aimed to obtain relevant data regarding the management of patients with mCRC and AEs after failure of first-line treatment in England and Wales. The response from each of the six physicians was collated and synthesised using summary statistics (median, mean, standard deviation). In the manufacturer's base case, the mean resource use of adverse events was calculated from the KOL survey to estimate the costs of AEs. 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. In a sensitivity analysis conducted by the manufacturer, applying the mean of the KOLs' responses to the model led to a considerable increase of the ICER (from £36,294 to £41,222 per QALY). The manufacturer did not comment on this result. It is unclear in this case whether the median is a better estimate than the mean (i.e. whether the skew reflects the distribution of resource use that might be expected in practice or whether this simply reflects variation between the KOLs and the different experience and understanding that they may have), but the model is clearly sensitive to the parameter. Given the magnitude of the effect on cost-effectiveness, the ERG believes that the uncertainty at this level should not be disregarded. Further exploration of this uncertainty area, is discussed in section 6.

5.2.8.5 Adverse events

The costs associated with adverse events from second-line treatment of mCRC included in the manufacturer's economic model, as well as the respective unit costs, are shown in Table 5.12. The costs for diarrhoea and febrile neutropaenia were based on NHS reference costs [52, 55]. All other estimates were calculated by combining resource use estimates collected in the physician survey and unit costs. The responses from the six KOL were synthesised by calculating the mean of all responses before application of unit costs. Resource use categories included to estimate costs of AEs treatment included: percentage of hospitalisations, percentage of outpatient attendances (specified as either GP or oncologist visits), routine tests and medications. Unit costs were not described by the manufacturer in the MS, but could be found in the model. Despite being referenced in the model as being taken from NHS reference costs 2011-12, unit costs for hospitalisations refer to the financial year 2010-11[53]. This constitutes an inconsistency, as NHS reference costs 2011-12[52] are used throughout the MS, and is likely to have underestimated overall costs. However, the ERG does not anticipate that this will change results significantly, as updating unit costs on hospitalisations for the 2011-12 financial year only increases the base case ICER to £36,455 per QALY.

Table 5.12 List of adverse events included in the model – adapted from the MS (Table B37)

Adverse event	Estimated Cost (£)	Unit costs for inpatient admissions	
		(£)	HRG code and description
Diarrhoea	475	475[55]	FZ45C- Non-Malignant Large Intestinal Disorders, with length of stay 1 day or less (all HRG data)
Hypertension	759	1,668	EB04I - Hypertension without CC
Fatigue	90	987	WA19Y- Abnormal findings without diagnosis without CC
Stomatitis	929	1,664	WA09Y - Other non-viral infection without CC
Abdominal pain	1,066	1,133	PA29Z - Abdominal pain
Vomiting	360	469	FZ43C -Non-Malignant Stomach or Duodenum Disorders with length of stay 1 day or less
Dehydration	672	955	KC05F - Fluid and Electrolyte Disorders 69 years and under without CC
Decreased appetite	0	987	WA19Y - Abnormal findings without diagnosis without CC
Nausea	310	469	FZ43C - Non-Malignant Stomach or Duodenum Disorders with length of stay 1 day or less
Neuropathy peripheral	288	1,331	AA26B - Muscular, Balance, Cranial or Peripheral Nerve disorders; Epilepsy; Head Injury
Infections	328	507	LA04G - Kidney or Urinary Tract Infections with length of stay less 1 day or less
Hand-foot syndrome	338	1,392	PA35B - Skin Disorders without CC
GI perforation	469	469	FZ43C - Non-Malignant Stomach or Duodenum Disorders with length of stay 1 day or less
Haemorrhage	576	576	FZ45C - Non-Malignant Large Intestinal Disorders with length of stay 1 day or less
Febrile neutropaenia	5,675	5,675[52]	PA45Z - Febrile Neutropaenia with Malignancy
Neutropaenia	1,586	4,227	PA45Z - Febrile Neutropaenia with Malignancy
Asthenia	263	987	WA19Y - Abnormal findings without diagnosis without CC

GI, gastrointestinal; HRG, Healthcare Resource Groups.

Unit cost values applied to the KOL survey resource use estimates refer to the national average cost for NHS Trusts and Primary Care Trusts combined Non-Elective Inpatient (Long Stay) HRG Data [53]

For some of the adverse events, there are also concerns regarding the face validity of the selection of unit costs for hospitalisations, and of the resulting estimated costs. It is arguable whether adverse events such as gastrointestinal perforation, should be considered to belong to the same HRG code as vomiting and nausea. It is also unclear the rationale for applying directly the unit cost of diarrhoea (£475) in the model, when there was an estimated cost from the physician survey (£667). Furthermore, the unit cost assigned to diarrhoea corresponds to an activity weighted average of all

HRG data from NHS reference costs 2011-12, whereas all other unit costs (with the exception of febrile neutropaenia) correspond to non-elective inpatient long-stay HRG data from NHS reference costs 2010-11[53]. The correspondent non-elective inpatient long-stay HRG code for 2011-12 is attached to a unit cost of £805. Another inconsistency was detected for the unit cost attached to febrile neutropaenia and neutropaenia. Despite both AEs sharing a common HRG code in the economic model, the unit cost estimates were taken from different sources (PbR tariff 2011-12 for febrile neutropaenia[55], and NHS reference costs 2011-12 for neutropaenia [52]).

All the detected inconsistencies impact to some extent on the intervention cost-effectiveness. Table 5.13 illustrates the relative impact on the ICER of the base case of solving some of these inconsistencies.

Table 5.13 Impact on cost-effectiveness estimates of correcting AEs costing inconsistencies

Adverse event	**Corrections**	**ICER (per QALY)**
-	None ; base case	£36,294
1. Diarrhoea	Estimated cost from KOL survey - £667	£36,408
2. Diarrhoea	Unit cost [52]- £805	£36,490
3. Neutropaenia	Unit cost[55] – £5,675	£36,413
4. Diarrhoea and neutropaenia	Combined 1. and 3.	£36,527
5. Diarrhoea and neutropaenia	Combined 2. and 3.	£36,609
KOL, Key Opinion Leader		

The ERG did not change these inconsistencies in the model, as the impact was considered minor. Nevertheless, it is important to highlight that all corrections would have led to an increase of the estimated ICER.

Another issue regarding the estimation of AEs costs that the ERG would like to highlight refers to the length of treatment. One of the outputs of the KOL survey is the length of inpatient stay associated with AEs treatment. However, this estimate was not used to estimate costs. For each AE, the unit cost of the correspondent HRG code was implicitly assumed to reflect the duration of inpatient stay. This could be reasonable if the estimated length of inpatient stay was similar to the average length of stay for each HRG code. The manufacturer did not state, nor justify, the assumption.

Adverse event costs per patient were included in the model by multiplying a weighted average of costs by the probability of incurring any of the included AEs at each cycle (see section 5.2.7 for details on AEs probability), and applying this to all patients in second-line treatment state. Costs of each adverse event were an average of the estimated costs in Table 5.12 weighted by the number of adverse events divided by total number of adverse events.

5.2.8.6 Unit cost inconsistencies

The ERG found additional inconsistencies regarding unit costs applied in the model. The "BNF, 2012" was referenced in the submission as the source for drug acquisition unit costs. However, the BNF detailed in the references section corresponds to the 62nd edition, 2011[38]. The ERG verified the unit costs for the individual drug components in FOLFIRI, and they correspond to those in the most recent BNF edition[56]. The unit costs of treatment and of other drugs (used in the management of mCRC and further lines of treatment) were not verified, given the great diversity of drugs included in the model and the limited time available to do the verification. These drug acquisition unit costs detailed in the model were referenced to the BNF 62[38].

Another inconsistency was detected on the unit cost selected to inform the cost of intensive care unit hospitalisations (included in management costs). The cited unit costs referred to adult inpatient attendances to mental health intensive care services[54]. No justification was given for the selection of a mental health services specific unit cost. The ERG estimated a weighted average of adult critical care unit HRGs from the NHS reference costs [52] that was considerably higher (£1,188.94 per bed day). Nevertheless, the economic model is not responsive to changes in this parameter (ICER remains £36,294 per QALY), so no correction was made by the ERG in Section 6.

5.2.9 Discounting

Both costs and benefits were discounted at an annual rate of 3.5%, as per the NICE reference case. The unit costs in the MS are stated to refer to the 2011-2012 financial year. Nevertheless, the ERG identified a number of unit costs at 2010-2011 prices, and for which no cost inflation was applied.

Superseded – see erratum

5.2.10 Cost effectiveness results

5.2.10.1Base case

In the manufacturer's base case costs and benefits were discounted at 3.5% per annum. The analysis time-horizon was 15 years. Survival functions fitted to each treatment group separately for PFS and OS estimated using the Weibull fitted to the IRC assessed PFS curve and log-logistic functions respectively. The cost of aflibercept and irinotecan assumed that unused drug in opened vials is discarded (no step-down to the nearest number of whole vials was assumed); the cost of wasted leucovorin and 5-FU was assumed to be negligible and similar for both treatment groups. Management costs were estimated based on the assumption that median estimates from the KOL survey were reflective of resource use in the UK routine clinical practice. Acquisition costs of drugs for second-line of treatment of mCRC are assumed to be incurred by the totality of each treatment cohort for a fixed number of cycles (correspondent to mean overall duration of combination treatment from the VELOUR trial). The cost-effectiveness results for the base case are reported in table 5.14, and refer to the overall population.

Table 5.14 Base case cost-effectiveness results - adapted from the MS (Table B44)

Technologies	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER (£) incremental (QALYs)
FOLFIRI	[redacted]	[redacted]	[redacted]	—	—	—	—
Aflibercept + FOLFIRI	[redacted]	[redacted]	[redacted]	8,816	0.3378	0.2429	36,294
LYG, Life Years gained.							

A similar set of assumptions appears to have been made for the base cases for each of the subgroups in the MS (a pre-specified liver metastases only subgroup and a *post-hoc* subgroup of patients that excluded those patients who adjuvant chemotherapy). However, different assumptions were made regarding PFS. For each subgroup, empirical PFS data from the VELOUR trial was used to model PFS, i.e. IRC assessed Kaplan-Meier curves specific to the subgroup (and treatment arm) were used instead of parametric curves. The manufacturer justified this based on a lack of acceptable fit of the parametric functions and the completeness of the observed data. No detail was provided for what the manufacturer considered an acceptable fit and whether it included visual fit and statistical fit (or both). The assumptions for the subgroup specific base cases were not described in the MS, but could be deduced from tables 10-24 and 10-24 (Appendix 16 of the MS). It is worth mentioning that other data specific to the subgroups used in the model included: OS empirical data; treatment

discontinuation curves; BSA and weight distributions (used for wastage calculations); mean dosage per administration and total number of administrations of each component of second-line of treatment; mean overall duration of combination treatment; and number of AEs observed. Cost-effectiveness results are show in Table 5.15.

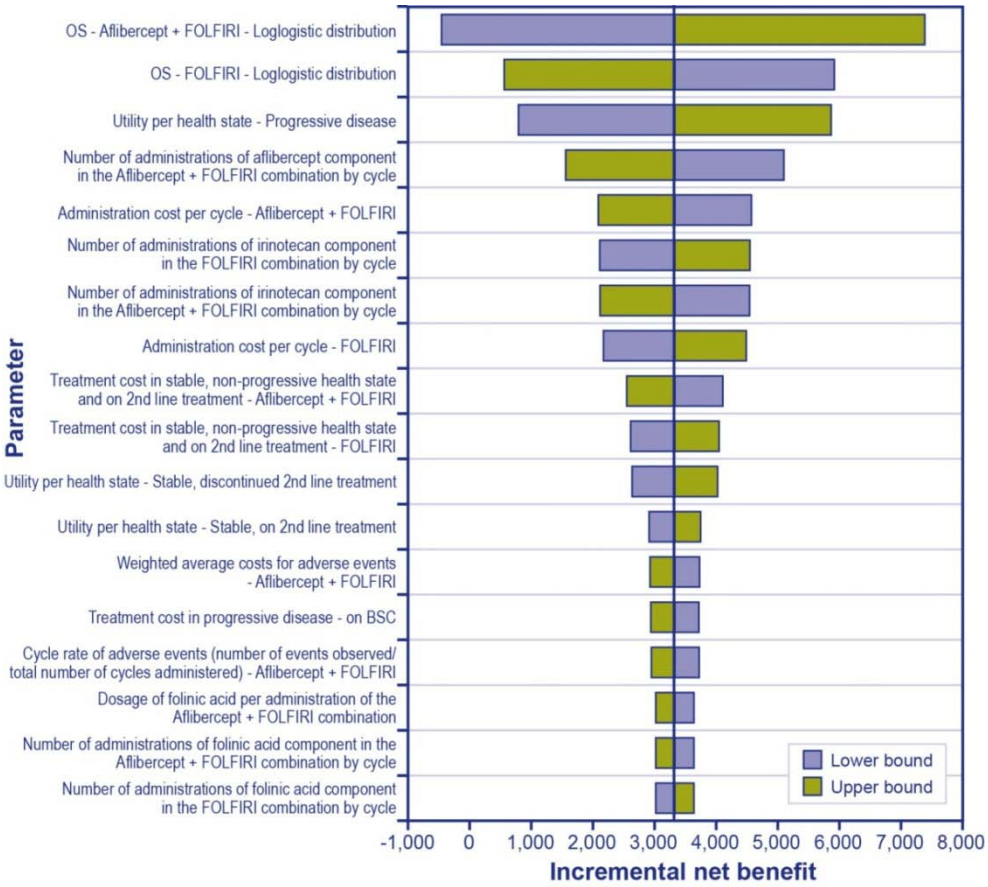
Table 5.15 Base case cost-effectiveness results - adapted from the MS (Table B46)

Technologies	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER (£) incremental (QALYs)
Liver metastases only subgroup							
FOLFIRI	[redacted]	[redacted]	[redacted]	—	—	—	—
Aflibercept + FOLFIRI	[redacted]	[redacted]	[redacted]	10,974	0.4955	0.3601	30,474
ITT population excluding patients who had relapsed on oxaliplatin-based adjuvant therapy							
FOLFIRI	[redacted]	[redacted]	[redacted]	—	—	—	—
Aflibercept + FOLFIRI	[redacted]	[redacted]	[redacted]	8,573	0.3668	0.2639	32,480
LYG, Life Years gained.							

5.2.10.2Sensitivity analyses

The manufacturer performed a series of one-way deterministic sensitivity analyses (DSA) to assess the impact of uncertainty around input variables on the net monetary benefit (NMB) at cost-effectiveness threshold of £50,000 per QALY. The selection of this particular threshold was not explicitly stated in the MS, but could be verified within the model. All model parameters were stated to have been varied across a range from [redacted] below to [redacted] above the initial base case point estimate. The results are presented in Figure 5.2 for the overall population. The results for the two subgroups can be found on Appendix 16 of the MS (Figures 10-13 and 10-16).

Figure 5.2 Manufacturer's deterministic sensitivity analyses results - overall patient population (Figure B-19, of the MS)



The MS also included a series of scenario analyses that were performed to check the robustness of the model to alternative structural assumptions. The main analyses are reported in Table 5.16 (equivalent results for the two subgroups are shown on tables 10-24 and 10-25 in Appendix 16 of the MS).

Table 5.16 Results of the scenario analyses (from Table B45, of the MS)

Parameter	Base case	Scenario	Incremental costs[a]	Incremental QALYs[a]	ICER
Base case results			8,816	0.2429	36,294
PFS function	Weibull	Log-logistic	8,562	0.2451	34,935
		Log-normal	8,594	0.2451	35,068
		Kaplan-Meier data	8,612	0.2429	35,459
		Weibull, progression determined by investigator assessment	8,802	0.2417	36,410
OS function	Log-logistic	Weibull	8,246	0.1656	49,805
		Log-normal	9,130	0.2702	33,788
Drug wastage	Unused drug in opened vials discarded[b]	No drug wastage (price per mg applied to mean dose)	[redacted]	[redacted]	[redacted]
		5% step-down scenario[a]	[redacted]	[redacted]	[redacted]
Management resource use from KOL survey	Median of KOL responses	Mean of KOL responses	10,013	0.2429	41,222
Time horizon	15 years	25 years	8,938	0.2587	34,549

5-FU, 5-fluorouracil; FOLFIRI, 5 fluorouracil/leucovorin/irinotecan; ICER, incremental cost-effectiveness ratio; KOL, key opinion leader, OS, overall survival; PFS, progression-free survival; QALY, quality-adjusted life-year.

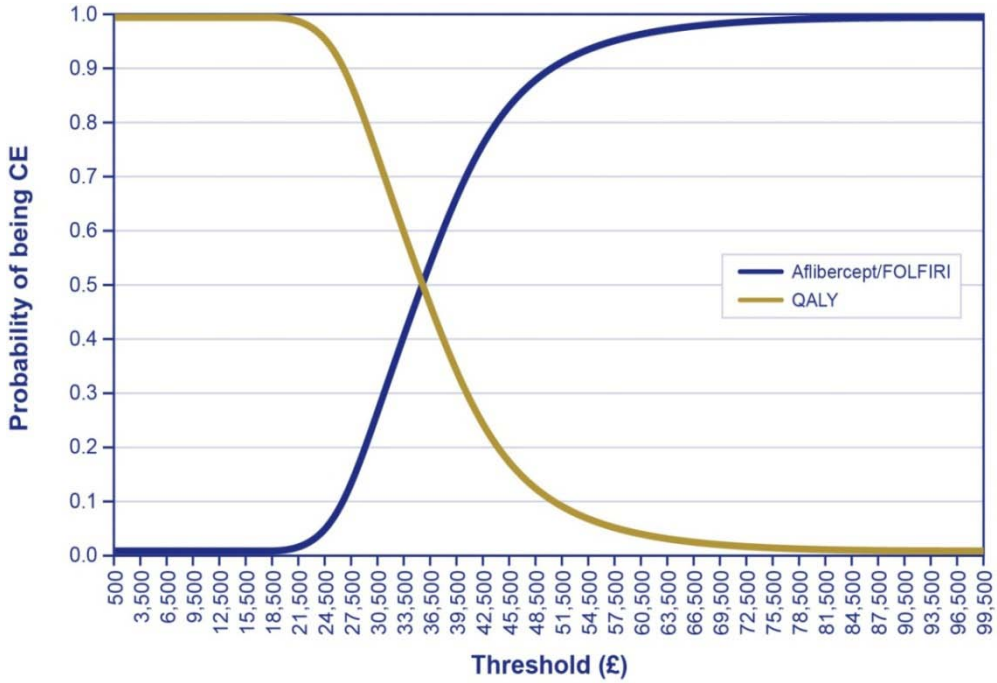
[a] Aflibercept + FOLFIRI vs. FOLFIRI.

[b] The cost associated with wasted leucovorin and 5-FU was assumed to be negligible and similar for both treatments and was excluded from the analysis.

Probabilistic sensitivity analysis (PSA) was performed to characterise uncertainty surrounding parameters on the model outputs. The manufacturer's model was set to run the PSA for 5,000 iterations. The cost-effectiveness acceptability curves (CEACs) for the base case of the overall population analysis was derived from the PSA and is shown on Figure 5.3. CEACs for the subgroup analyses are reported in Appendix 16 of the MS (Figures 10-15 and10-18). PSA results were also represented graphically on the cost-effectiveness plane (Figure B-20, MS; Figure 10-14 and 10-17, Appendix 16), and probabilities of cost effectiveness at different thresholds were reported. The

probability of cost-effectiveness was 89.5% at a cost-effectiveness threshold of £50,000 per QALY decreasing to 22% when the threshold was reduced to £30,000 per QALY.

Figure 5.3 Cost-effectiveness acceptability curves (Figure B-21, of the MS)



The manufacturer assessment of the univariate DSA concluded that the ICER estimate was most sensitive to the parametric assumptions regarding the distribution of overall survival, utility weight for the progressive disease health state and number of administrations of second-line treatment drugs.

[redacted] Another issue that the ERG would like to highlight is that, since results were presented in terms of NMB, they assume an implicit valuation of health benefits that may not be aligned with that of decision maker, i.e. NICE.

Superseded – see erratum

[redacted]

Regarding the scenario analyses, the manufacturer did not present an interpretation of the range of ICERs obtained in the analyses. The scenario analyses suggest that the model is sensitive to assumptions on OS and of no wastage. The latter is considered of less importance by the ERG, as it cannot be assumed that this will happen in a routine clinical practice setting. As the ERG has mentioned earlier in the report, the model is also sensitive to the selection of median or mean estimate of resource use from the KOL survey. It should also be noted that the manufacturer did not present any results regarding time horizons shorter than that of the base case, but results for longer time horizons are more favourable for aflibercept. Finally, the ERG believes that for completeness, the manufacturer should have reported cost-effectiveness results for the proportional hazards OS model equivalent to their base case, i.e. where OS empirical data is extrapolated by fitting dependent log-logistic curves with treatment as a predictor. When the ERG used this alternative extrapolation model, the ICER increased to £58,784 per QALY, with the difference being driven mainly by a 43.8% reduction in QALY gain compared to the manufacturer's base case.

The actual cost-effectiveness results reported in the MS are based on the deterministic analyses, rather than on the PSA as recommended in the NICE reference case.[40] Given the non-linearity that is characteristic of Markov models, probabilistic results are considered the best estimates of cost-effectiveness [20, 40]. The degree of non-linearity in the model will then determine whether the

deterministic and probabilistic ICERs differ, as well as the magnitude of this difference. Another important issue is the ICER stability to the number of iterations used in the PSA simulation. It is important that enough iterations are run, to ensure that the resulting ICER reflects the full sampling from the parameters probability distributions. The ERG performed a series of five runs of the manufacturer's model to assess the degree of non-linearity and to verify the model's stability at 5,000 iterations. The ICER estimates based on PSA did not differ considerably from the deterministic base case estimate (£36,294 per QALY), with percentage difference between each model run ranging from -2.678% to -2.025%. The PSA derived ICERs were also very stable between each 5,000 iterations model run (less than 1% variation between each model run and average ICER for the 5 runs). There is however another issue that may have led to uncertainty not being fully reflected in the PSA estimates. The standard errors on some model parameters, namely administration costs per cycle, weighted average costs and disutilities for AEs, dosage per administration for the irinotecan component (in both treatment arms), proportion of patients on further lines of treatment (pre and post-progression), and aggregated drug acquisition costs, were not estimated standard errors for those parameters. These standard errors were calculated assuming that the confidence interval comprised values 30% above and below the mean estimate for each parameter. This is an arbitrary variation of the parameter that is unlikely to be reflective of its actual uncertainty.

5.2.11 Model validation

The clinical validation of the model structure, parametric survival analyses, and parameter estimates was performed by an advisory board. Three advisory board meetings were held, one during the model-design phase and two after completion of the model. The composition of the advisory board and the criteria followed to select its members is not described in the MS.

The face validity was checked by a peer review board, described as a review committee consisting of internal senior scientists with extensive modelling and disease experience.

The manufacturer states that quality control procedures were in place, and included extreme-value sensitivity analyses and validation of the logical structure of the model, mathematical formulas and sequences of calculations, and the values of numbers supplied as model inputs. The model was said to be reviewed by a peer reviewer not involved with the original programming.

An independent modelling group performed further validation of the model. All input data were verified with original sources; programming was checked to verify that formulae were applied correctly; range and logic checks were performed using extreme values; and key calculations were independently reproduced.

Although the manufacturer appears to have followed a comprehensive internal and external validation of their results, we remain concerned regarding the external validity (face-validity) of the projected survival gains and the assumptions made concerning HRQoL, and their impact on the overall cost-effectiveness results. Furthermore, in the course of our review, the ERG identified a programming error in the model which has been discussed in Section 5.2.7.3, and is addressed in Section 6.

5.3 Exploratory and sensitivity analyses undertaken by the ERG

Importantly, due to the complex interplay between many of the assumptions highlighted in this chapter, it is difficult to conclude that individual parameter estimates or assumptions are in themselves likely to lead to optimistic or conservative estimates of the ICER. The ERG conducted a large number of exploratory analyses in order to investigate the impact of these factors and their interactions. The ERG considers that many of the key issues identified in this section need to be considered jointly, in order to more fully characterise the nature of the uncertainty and to provide a more appropriate basis for informing the ICER estimates. Hence, the ERG has selected the most important issues to present within the report. These key issues include the extrapolation of OS and the HRQoL in PD state. Consequently, much of Section 6 is devoted to a joint assessment of these key issues, and the impact on the ICER based on additional exploratory analyses undertaken by the ERG.

5.4 Conclusions of the cost effectiveness section

The manufacturer's state that the data from the mCRC Utilities study, mCRC UK Resource Use study and VELOUR trial are representative of the UK patients in routine clinical practice. They claim that the VELOUR trial results on OS and PFS are robust and require little extrapolation and suggest the economic results are robust to sensitivity analyses and scenario analyses performed. The manufacturer reported a base case ICER for aflibercept + FOLFIRI vs. FOLFIRI alone of £36,294 per QALY.

The ERG identified the extrapolation of the OS data as a key driver of the cost-effectiveness results By fitting independent parametric functions to the OS data (as opposed to assuming proportional hazards) the manufacturer's base case implicitly assumes that the relative treatment effect of aflibercept on OS continues to improve over time, even after second-line treatment discontinuation and disease progression. The ERG considers the assumptions on OS, and how they translate in terms of the assumed duration and magnitude of treatment effect, to be a major area of uncertainty, and that further analyses are required to explore it. The ERG also considers the utility values applied by the manufacturer to be another key source of uncertainty. The utility value associated with the progressive disease state was higher than has been used in previous analyses and is unlikely to appropriately reflect HRQoL changes over the entire post-progression period. Furthermore, the utility values applied in the model did not explicitly adjust for age; neither taking into account the higher ages of patients in routine clinical practice (compared to the younger trial population) nor that the patient population itself will continue to age as the model time horizon increases. Consequently the utility values used by the manufacturer are potentially optimistic towards the cost-effectiveness of aflibercept.

6 IMPACT ON THE ICER OF ADDITIONAL CLINICAL AND ECONOMIC ANALYSES UNDERTAKEN BY THE ERG

6.1 Overview

This section presents the results of the additional cost-effectiveness analyses undertaken by the ERG. Following on from the discussion in Section 5, we begin by correcting the manufacturer's model so that acquisition and administration costs of second-line treatment are applied to all patients in the second-line treatment state, instead of assuming a fixed number of cycles. The duration of adverse events is corrected in the model to avoid duplicating the adjustment for cycle length when calculating total undiscounted QALYs in the model. The model is also adapted to include a decrement in the utility for aging. The impact of each change on the ICER is highlighted alongside their combined impact.

We begin our additional analyses by demonstrating the sensitivity of the cost-effectiveness results to the duration of the treatment effect for OS, by highlighting the relationship between the ICER results and the model time horizon. We then explore the robustness of the cost-effectiveness results to alternative assumptions concerning the treatment effect on OS. Additional uncertainties highlighted in Section 5 regarding utility estimates, with particular emphasis on the estimates of post-progression utility, are also explored. Additionally, the ERG will demonstrate the impact of considering the additional costs associated with delivering the aflibercept infusion and explore the effect of using the mean versus the median resource use as provided by the KOL survey.

Finally, we will define an alternative ERG base case analysis which we consider represents a more appropriate set of assumptions given the current evidence. We conclude by re-examining the impact on cost-effectiveness of the duration and magnitude of the treatment effect when combined with the alternative ERG base case.

6.2 ERG corrections and adjustments to the manufacturer's base case model

Acquisition and administration costs of second-line treatments were included in the model for a fixed number of cycles, which corresponded to the mean number of cycles of combination treatment administered to each treatment arm in the VELOUR trial. As discussed in section 5.2.8.1, the acquisition and administration 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. Therefore, the ERG applies the administration and acquisition costs of aflibercept + FOLFIRI and FOLFIRI to all patients in the second-line treatment substate. This correction increases the base case ICER to £37,539 per QALY.

It was also previously noted in section 5.2.7.3 that the AE related disutility was adjusted twice for cycle length. The ERG corrected the model by removing the cycle adjustment for AE duration, which increased the base case ICER to £37,834 per QALY.

To account for the impact of age on the HRQoL an age dependent utility decrement was applied to the model. The utility decrement was first applied to a starting population of 60 years old representing the average age of the VELOUR trial. The ERG then assumed that the patients entered the model at the age of 70 years old, which was considered to be the median age at diagnosis in mCRC in previous NICE guidance [10]. To implement these changes, the ERG tracked the age of each patient by assuming a starting age in the first cycle of the model and then increasing the age at each cycle by the cycle length. A utility decrement was calculated by subtracting the utility from each age interval by the utility from the 60-65 years old interval, which represented the age in the manufacturer's Utility Study (Table 6.1). The decrement was converted to a QALY decrement by adjusting for the cycle length. The age dependent QALY decrement was multiplied by the proportion of patients in each alive health state and subtracted from the QALY at each cycle of the model. Utilities were estimated from the Health Survey for England [49] and are reported below in table 6.1. Assuming a starting age of 60 and accounting for the impact of age on HRQoL increased the ICER to £36,456 per QALY. Assuming a starting age of 70 and accounting for the impact of age on HRQoL increased the ICER to £38,500 per QALY.

Table 6.1 Utility decrement associated with ageing (unadjusted for cycle length)

Age interval (years)	Utility[49]	Utility decrement
60-65	0.8072	0
65-70	0.8041	-0.0031
70-75	0.779	-0.0282
75-80	0.7533	-0.0539
80-85	0.6985	-0.1087
85+	0.6497	-0.1575

The combination of all corrections and utility adjustments made to the model yields an ICER of £41,653 per QALY.

Superseded – see erratum

6.3 Additional ERG sensitivity analyses

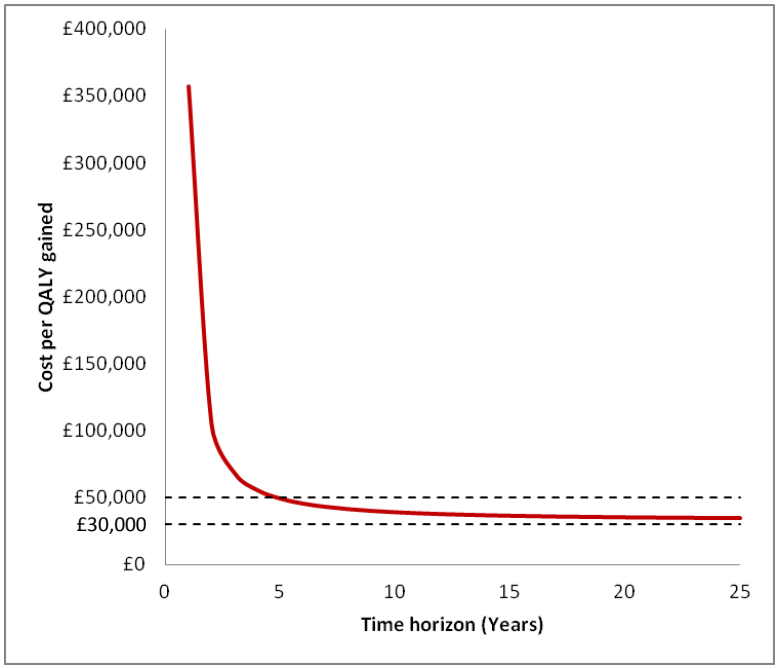
The following sections report the results of the additional sensitivity analyses performed by the ERG to explore areas of uncertainty and issues that were highlighted in Section 5.

6.3.1 Treatment effect duration

The OS extrapolation method has been identified in Section 5 as the major driver of cost-effectiveness in the manufacturer's economic model. The model is very sensitive to the selection of extrapolation model making it a major area of uncertainty. 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. However, equally the results from the manufacturer's piecewise model clearly showed that the uncertainty surrounding these estimates also increased over time (with a decreasing number of patients and wider CIs). The PFS curves in the trial exhibited a similar initially divergent treatment effect pattern, but converged by 12 months. All these factors suggest that the sustained increase throughout time in OS gains assumed for aflibercept + FOLFIRI, is very uncertain in terms of magnitude and duration, and appears optimistic towards the cost-effectiveness of the intervention.

The ERG first explored the sensitivity of the cost-effectiveness results to the time horizon of the model to provide an indication of how influential the period of extrapolation is to these results. The results of the analysis are illustrated by Figure 6.1 that plots the ICERs as a function of the time horizon.

Figure 6.1 The ICER as a function of time horizon



The analysis shows that in the manufacturer's base case analysis aflibercept + FOLFIRI is only likely to become cost-effective at a threshold of £50,000 per QALY, if the treatment effect estimates assumed by the manufacturer continue for at least 5 years. However, within the range of the time horizons, aflibercept is never likely to be cost-effective at a threshold of £30,000 per QALY. Although this analysis clearly indicates the importance of the model time horizon for the ICER estimates, it should also be appreciated that this approach does not provide an appropriate basis for informing decision making. By restricting the horizon in this manner, the model is making an extreme assumption that all patients (in both arms) immediately die at the specified time horizon. This assumption is often termed "stop and drop" since it does not attempt to account for the longer term consequences (i.e. additional life years gained and QALYs) that should be attributed to the differential mortality rates that have been accumulated up until to the specified time horizon. That is, even if additional treatment benefits are not incorporated beyond at a particular time point, restricting the model time horizon to match this would likely provide an underestimate of the overall survival gains and result in an overly conservative estimate of the ICER. Instead, where there exists uncertainty concerning both the magnitude and duration of OS benefits, a more appropriate basis 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.

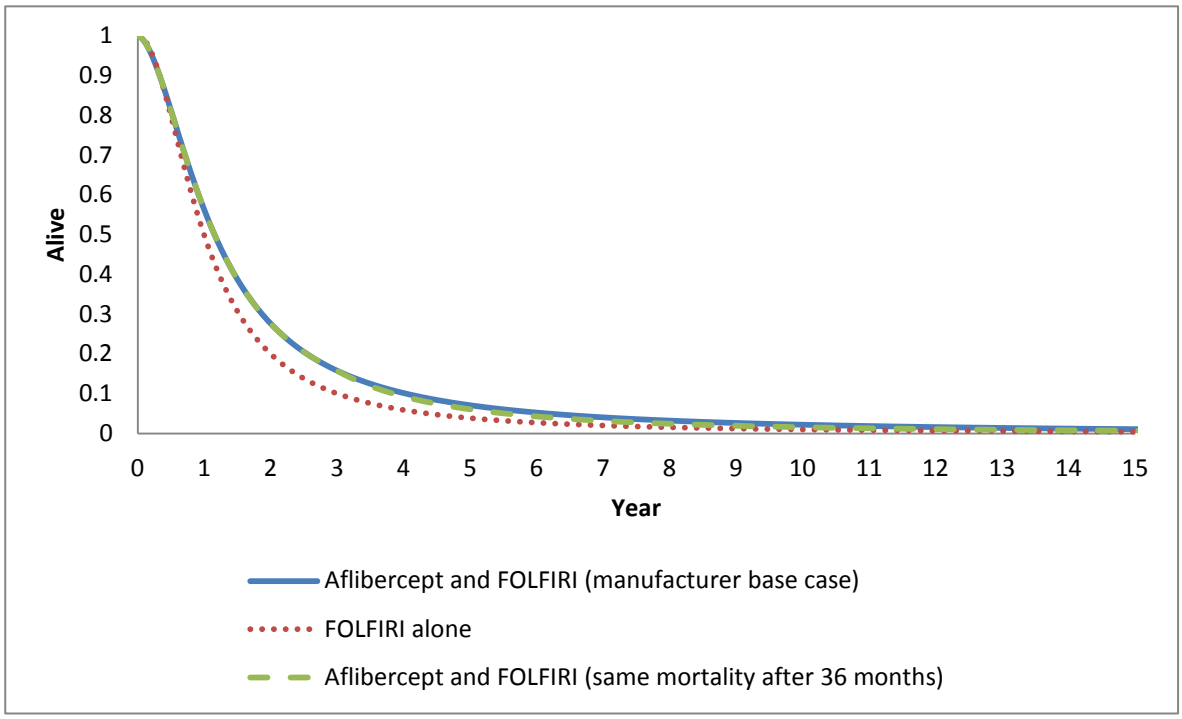
Superseded – see erratum

Since the treatment effect may not continue increasing for the entire duration of the model time horizon, the ERG tested alternative assumptions on the duration of the treatment at two time points previously identified as being relevant to the economic analysis, 30 and 36 months. The ERG thus investigated a series of scenarios to explore the robustness of the ICER estimates to alternative assumptions concerning the magnitude and duration of the effect on OS. Different methods for adjusting the duration and magnitude of treatment effect were assumed, i) risk of death are the same between treatment arms beyond 30 or 36 months, 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 converge at an earlier point). The ERG considers that these scenarios represent important extensions to the manufacturer's base case approach and together the combined set of analyses more appropriately reflect the full range of scenarios outlined in the NICE methods guide[40]} required to characterise uncertainty surrounding the extrapolation approaches. Each scenario is described below.

Previously the ERG has reported the ICER from the manufacturer's model using OS curves assuming proportional hazards, £58,784 per QALY. Although the ICER is less favourable when the proportional hazard model is applied, the ERG considers that even this scenario may be relatively optimistic, particularly given the convergence reported for the PFS curves.

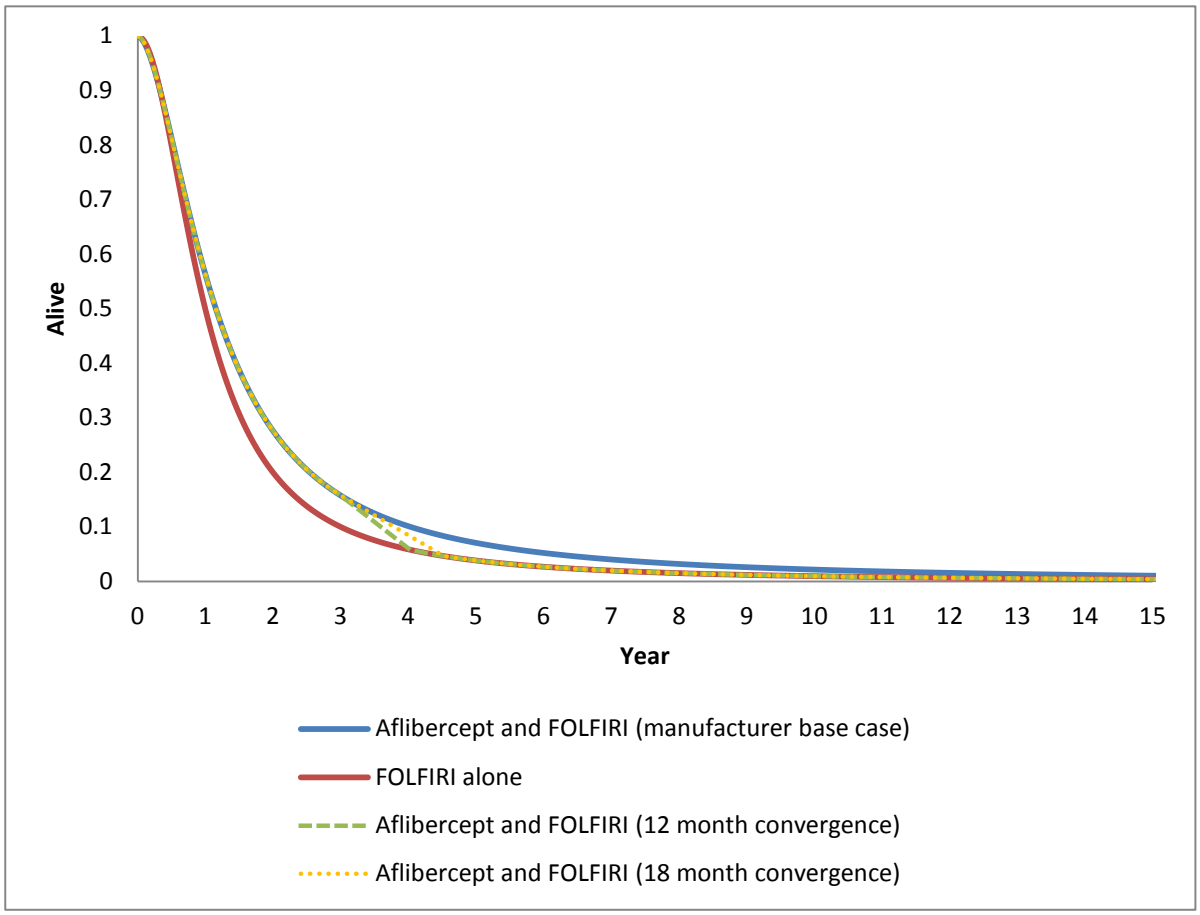
The ERG first tested the extrapolation of the OS assuming the difference in the treatment effect increases (as in the manufacturer's base case) until 30 or 36 months after which the risk of death in both arms is assumed to be the same (although applied to a different number of patients), i.e. there is no continued treatment effect. This scenario assumes no difference in the hazard rates of each treatment for OS beyond 30 or 36 months. The proportion of patients alive at each time point and treatment arm is shown in Figure 6.2 for the manufacturer's base case and for the ERG scenario analyses when the duration of treatment effect is 36 months.

Figure 6.2 Proportion of patients alive over time – ERG scenario assuming no continued treatment effect after 36 months.



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. This is modelled by assuming that the OS curves converge between 12 to 18 months after the observed follow-up. In order to do so, the ERG modelled scenarios where the OS curve for aflibercept + FOLFIRI began converging towards the FOLFIRI curve at 30 or 36 months, with full convergence after an additional 12 or 18 months. This is illustrated in Figure 6.3, where OS curves for convergence starting at 36 months are shown alongside the manufacturer's base case. For each time point (30 or 36 months), the scenarios with convergence of OS curves over 12 months assume that the magnitude of the aflibercept treatment effect is declining at a faster rate than those where convergence occurs within 18 months, and should therefore result in more conservative ICERs. At the time point at which the curves crossed the ERG assumed the mortality rate was identical in both arms for the remainder of the model time horizon.

Figure 6.3 Proportion of patients alive over time – ERG scenarios assuming that the OS curves converge over 12 or 18 month period.



The cost-effectiveness results for the 6 resulting scenarios are presented in Table 6.2. In the scenarios where treatment effect is assumed to last 36 months the ICER estimates for aflibercept + FOLFIRI range from £42,718 to £59,666 per QALY. For the scenarios where the additional assumptions related to treatment effect magnitude and duration are applied at 30 months, the ICERs range between £45,570 and £66,377 per QALY. It is not surprising that the ICERs are lower when the treatment effect is assumed to last longer, i.e. the scenarios assuming treatment duration of 36 months versus 30 months.

Scenarios that assumed convergence of the OS curves over a 12 month period resulted in higher ICERs than those scenarios where convergence occurs over an 18 month period. The ICERs are lower than £50,000 per QALY for those scenarios for which there is no continued treatment effect after 30 or 36 months.

Since these scenarios represent more conservative assumptions regarding the magnitude and duration of the treatment effect estimates for OS applied during the extrapolation period, all of these additional analyses produced less favourable ICERs compared to the manufacturer's base case approach. These scenarios are further explored jointly with the additional parameter assumptions considered within the revised ERG base case reported in section 6.4.

Table 6.2 Cost-effectiveness results for different assumptions on treatment effect duration

Scenario	Aflibercept + FOLFIRI		FOLFIRI		Inc. Costs (£)	Inc. QALYs	ICER (£/QALY)	% Change
	Mean costs (£)	Mean QALYs	Mean costs (£)	Mean QALYs				
Base case	[redacted]	[redacted]	[redacted]	[redacted]	8,816	0.2429	36,294	-
1.1. No continued effect after 30 months	[redacted]	[redacted]	[redacted]	[redacted]	8,358	0.1834	45,570	25.56
1.2. OS curves converge over 18 month period (months 30 to 48)	[redacted]	[redacted]	[redacted]	[redacted]	7,980	0.1343	59,428	63.74
1.3. OS curves converge over 12 month period (months 30 to 42)	[redacted]	[redacted]	[redacted]	[redacted]	7,858	0.1184	66,377	82.89
2.1 No continued effect after 36 months	[redacted]	[redacted]	[redacted]	[redacted]	8,473	0.1984	42,718	17.70
2.2. OS curves converge over 18 month period (months 36 to 54)	[redacted]	[redacted]	[redacted]	[redacted]	8,067	0.1456	55,424	52.71
2.3. OS curves converge over 12 month period (months 36 to 48)	[redacted]	[redacted]	[redacted]	[redacted]	7,976	0.1337	59,666	64.40

6.3.2 Post-progression utility

In section 5 the ERG has highlighted a number of concerns regarding the utility estimates used in the economic model. The utility estimates used by the manufacturer appear high, considering the age and health status of the population with mCRC in routine clinical practice. Indeed, the mCRC Utilities Study was conducted in patients younger (and possibly fitter) than mCRC patients treated in clinical practice in the UK population.

The ERG also considered that the utility estimate in PD state was unlikely to fully reflect the post-progression HRQoL in mCRC patients. The concerns around the utility estimate in the PD state relate not only to the uncertainty on the absolute value of the estimate used in the model ([redacted]), but also with the clinical plausibility of the utility in the state remaining constant over time. In previous NICE appraisals in mCRC [10 1, 43, 57, 58], the appropriateness of utility estimates in post-progression states have been questioned, as they were considered not to adequately reflect the decline in HRQoL that is expected to occur towards the terminal phase of the disease. The ERG considered the utility estimates from sources in the manufacturer's systematic review, and identified the utility estimates in Table 6.3 as providing a range of potentially relevant post-progression utility estimates.

Table 6.3 Utility estimates in Progressive Disease for mCRC

Health state	**Utility estimate**	**Source**
Progressive disease	0.60	ERG report [58]
	0.69	ERG report [1 10]
	0.21*	CG131 evidence review[6]
	0.29**	Best *et al.*, 2010[45]
* CRC patients only estimate for metastatic progressive disease, from Best *et al.*, 2010 ** Combined community and CRC patients estimate for metastatic progressive disease		

The ERG recognizes that the 0.21 and 0.29 estimates are unlikely to appropriately characterise the entire post-progression period but may provide an indication of the magnitude of change that might arise closer to the terminal phase of the PD state. Conversely, the mCRC Utilities Study estimate of [redacted], as well as the previously used 0.69 in a previous HTA,[1, 10] may be an overestimate of the 'average' utility over the entire post-progression survival period. Despite the utility estimate used in an ERG revised analyses for a previous HTA[58] having been calculated by applying an arbitrary reduction of 25% to the stable non progressive disease utility estimate, this value may well reflect a better compromise between the two extremes of the considered range. Thus, the ERG selected to use 0.60 as an estimate of utility for the PD state. The cost effectiveness results are reported in Table 6.4.

Table 6.4 Cost-effectiveness results for alternative assumption on PD utility

Scenario	aflibercept + FOLFIRI		FOLFIRI		Inc. Costs (£)	Inc. QALYs	ICER (£/QALY)	% Change
	Mean costs (£)	Mean QALYs	Mean costs (£)	Mean QALYs				
Base case PD utility = [redacted]	[redacted]	[redacted]	[redacted]	[redacted]	8,816	0.2429	36,294	-
3.1 PD utility = 0.21	[redacted]	[redacted]	[redacted]	[redacted]	8,816	0.1239	71,143	96.02
3.2 PD utility = 0.60	[redacted]	[redacted]	[redacted]	[redacted]	8,816	0.2171	40,608	11.89

As expected by lowering the utility of the PD state from [redacted] the mean QALYs decreased in both arms, but most importantly so did the incremental QALYs, resulting in a 12 to 96% higher ICER at £40,608- £71,143 per QALY.

6.3.3 Additional administration costs for aflibercept

The administration of aflibercept + FOLFIRI may require an additional hour of infusion time when compared to administering FOLFIRI alone. The manufacturer did not include in the analysis any additional costs due to this extra infusion time. As highlighted in Section 5.2.8, the ERG believes that this could have underestimated the costs of administration for the aflibercept + FOLFIRI treatment group. This cost was included in an exploratory analysis as an additional cost of £45 for those patients who had received aflibercept. The £45 estimate corresponds to the difference in cost between delivering a more complex and a simple parenteral chemotherapy at first attendance from the NHS reference costs.[52] Table 6.5 reports the HRG codes and respective unit costs used by the ERG to derive this value.

Table 6.5 Chemotherapy delivery unit costs (day case and regular day/night)[52]

HRG code	HRG Description	National Average Unit Cost
SB12Z	Deliver Simple Parenteral Chemotherapy at First Attendance	£203
SB13Z	Deliver more Complex Parenteral Chemotherapy at First Attendance	£248
HRG, Healthcare Resource Groups		

Additionally, the manufacturer's model did not include the preparation costs for the additional infusional treatment in the aflibercept + FOLFIRI arm. A pharmacy preparation cost of £15 per infusion was included by the ERG in the administration costs of aflibercept. This value was estimated by the ERG for a previous NICE HTA[1], and its use by the ERG assumes that the preparation time of aflibercept is similar to that of bevacizumab and cetuximab.

As expected the mean QALYs do not change but the mean costs of aflibercept + FOLFIRI increase (Table 6.6). Including the additional cost of pharmacy preparation and the cost of the additional administration time increases the ICER by 8% to £39,258 per QALY.

Table 6.6 Cost-effectiveness results for alternative assumptions on administration costs

Scenario	aflibercept + FOLFIRI		FOLFIRI		Inc. Costs (£)	Inc. QALYs	ICER (£/QALY)	% Change
	Mean costs (£)	Mean QALYs	Mean costs (£)	Mean QALYs				
Base case No additional administration costs	[redacted]	[redacted]	[redacted]	[redacted]	8,816	0.2429	36,294	-
4.1 aflibercept related costs = £15	[redacted]	[redacted]	[redacted]	[redacted]	8,996	0.2429	37,035	2.04
4.2 aflibercept related costs = £45	[redacted]	[redacted]	[redacted]	[redacted]	9,356	0.2429	38,517	6.13
4.3 aflibercept related costs = £60	[redacted]	[redacted]	[redacted]	[redacted]	9,536	0.2429	39,258	8.17

6.4 ERG alternative base case and OS scenarios

The results of the ERG additional analyses allowed some initial exploration of the model's robustness to alternative assumptions on treatment effectiveness, HRQoL and costs. It has been demonstrated how uncertainty on the assumptions made by the manufacturer are likely to individually impact on the cost-effectiveness of aflibercept + FOLFIRI. However, it is also important to consider the impact of these changes in combination. To investigate the impact on cost-effectiveness of the joint uncertainty associated with these assumptions, the ERG revisits the set of scenarios for alternative assumptions on treatment effect duration, discussed in section 6.3, using the ERG corrections described previously and applying the following additional assumptions:

1. Utility of PD state of 0.60;
2. An additional administration cost for aflibercept of £15;
3. Mean resource use from the KOL survey.

These estimates are used to define an alternative ERG base case analysis. This base case applies 0.60 as the estimate for PD as a compromise between the extremes reported in different studies as discussed earlier. The ERG also included an additional cost of £15 to the administration of aflibercept. Importantly, within the alternative ERG base case, the additional £45 cost of extra administration time that was considered plausible was not included based on the manufacturer's assertion that aflibercept will be administered at the same time as FOLFIRI suggesting that it will incur no additional nurse time. If this is not the case then the estimates provided will be an underestimate of the total costs of aflibercept and the ICER.

It was discussed in Section 5.2.8.4 the appropriateness of using the mean, instead of the median, as an estimate of resource use from the KOL survey. As discussed, mean values are generally the most appropriate estimates to inform parameters in cost-effectiveness analysis. Moreover, since the manufacturer applied mean resource use estimates from the KOL survey to estimate the cost of AEs management, it is more consistent to use mean estimates for the community care and social and personal care costs. Consequently, the mean values are applied throughout the resource use estimates for the ERG revised base case. However, since the KOL survey contained a small number of respondents (n=6) and estimates were considerably skewed, the ERG also presents an additional set of results based on the median resource use from the KOL survey in Appendix 1.

The OS extrapolation method has been identified in Section 5 as the major driver of cost-effectiveness in the manufacturer's economic model, but is also the most uncertain aspect of the model. By fitting separate parametric curves to each treatment arm, it was assumed in the MS that the treatment effect on OS will continue to increase over the extrapolation period. Given the high uncertainty surrounding the extrapolation of the OS data, the ERG explored the extrapolation scenarios tested in Section 6.3.1 using the ERG's alternative base case.

The revised ERG base case combined with the alternative OS extrapolation scenarios resulted in ICERs ranging from £47,965 to £92,089 per QALY (Table 6.7). Importantly, the ICER estimates across all these scenarios are consistently less favourable than the manufacturer's base case ICER. The magnitude of the difference between the manufacturer's base case ICER and those of the

alternative scenarios depends mostly on the assumptions on treatment effect duration. In the scenarios where aflibercept's treatment effect is assumed to continue to increase at least 36 months, the ICER estimates for aflibercept + FOLFIRI compared to FOLFIRI range from £56,197 to £83,605 per additional QALY. Applying this same treatment effect until 30 months, the ERG estimates ICERs to be between £59,743 and £92,089 per QALY. The assumption of no continued treatment effect is more favourable to aflibercept, than that of a declining treatment effect after the same time point and until OS curves convergence, with ICERs ranging between £56,197 and £66,506 per QALY in the case with no continued treatment effect, and between £ 71,635 and 92,089 per QALY for declining treatment effect. Scenarios where the decline of the treatment effect occurs within a shorter time period (12 vs. 18 months), i.e. with treatment effect declining at a higher rate, consistently yield higher ICERs. Furthermore, in the ERG's alternative analyses where utility of PD state was assumed lower (0.60 vs. [redacted]) than the base case manufacturer, the ICERs were all greater than £50,000 per QALY and ranged between £62,894 and £92,089 per QALY.

Model results based on the probabilistic sensitivity analysis for selected scenario analyses are reported in Table 6.8. The results of the probabilistic sensitivity analysis suggest that the deterministically derived results underestimate the ICER, by £5,000 to £14,000 per QALY in the scenarios undertaken. This suggests an increase on the degree of non-linearity of the model, and may be related with the manufacturer's assumptions on the standard errors of some parameters.

Further analyses using the median resource use from the KOL survey are reported in Appendix 1. The results in Appendix 1 were estimated by applying to the model the median resource use estimates of the survey to the ERG alternative base case and OS scenarios. By comparing the equivalent scenarios in Appendix 1 to those in Table 6.7 it can be concluded that the use of the median resource use consistently decreases the ICERs across scenarios (approximately £5,000 per QALY decrease in the ICERs for the alternative OS scenarios).

Table 6.7 Cost effectiveness results for the ERG alternative scenarios for the overall population, assuming mean (KOL survey) resource use

Model	Extrapolation Scenario	PD utility	Aflibercept + FOLFIRI		FOLFIRI		Inc. Costs (£)	Inc. QALYs	ICER (£/QALY)	% Change
			Mean costs (£)	Mean QALYs	Mean costs (£)	Mean QALYs				
MS Base case	Not applicable	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	8,816	0.2429	36,294	-
ERG base case	5.1 Manufacturer's base case independent log-logistic functions	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	10,509	0.2191	47,965	32.16
		0.600	[redacted]	[redacted]	[redacted]	[redacted]	10,509	0.1933	54,368	49.80
ERG base case	6.1 No continued effect after 30months	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	9,829	0.1645	59,743	64.61
		0.600	[redacted]	[redacted]	[redacted]	[redacted]	9,829	0.1478	66,506	83.24
	6.2 OS curves converge over 18 month period (months 30 to 48)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	9,155	0.11881	77,056	112.31
		0.600	[redacted]	[redacted]	[redacted]	[redacted]	9,155	0.1096	83,503	130.08
	6.3 OS curves converge over 12 month period (months 30 to 42)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	8,944	0.1039	86,056	137.11
		0.600	[redacted]	[redacted]	[redacted]	[redacted]	8,944	0.0971	92,089	153.73
	7.1 No continued effect after 36 months	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	10,033	0.1785	56,197	54.84
		0.600	[redacted]	[redacted]	[redacted]	[redacted]	10,033	0.1595	62,894	73.29
	7.2 OS curves converge over 18 month period	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	9,314	0.1300	71,635	97.38

Model	**Extrapolation Scenario**	**PD utility**	**Aflibercept + FOLFIRI**		**FOLFIRI**		**Inc. Costs (£)**	**Inc. QALYs**	**ICER (£/QALY)**	**% Change**
			Mean costs (£)	**Mean QALYs**	**Mean costs (£)**	**Mean QALYs**				
MS Base case	Not applicable	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	8,816	0.2429	36,294	-
	(months 36 to 54)	0.600	[redacted]	[redacted]	[redacted]	[redacted]	9,314	0.1191	78,226	115.53
	7.3 OS curves converge over 12 month period (months 36 to 48)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	9,152	0.1186	77,161	112.60
		0.600	[redacted]	[redacted]	[redacted]	[redacted]	9,152	0.1095	83,605	130.35

Table 6.8 Selected PSA cost effectiveness results for the ERG alternative scenarios overall population

	Scenario	Aflibercept + FOLFIRI		FOLFIRI		Inc. Costs (£)	Inc. QALYs	ICER (£/QALY)	% Change
		Mean costs (£)	Mean QALYs	Mean costs (£)	Mean QALYs				
Base case	Not applicable	[redacted]	[redacted]	[redacted]	[redacted]	8,816	0.2429	36,294	-
ERG base case	Not applicable*	[redacted]	[redacted]	[redacted]	[redacted]	11,821	0.19	61,446	69.30
Treatment effect duration	6.1 No continued effect after 30months *	[redacted]	[redacted]	[redacted]	[redacted]	11,387	0.16	71,732	97.64
	6.3 OS curves converge over 12 month period (months 30 to 42)*	[redacted]	[redacted]	[redacted]	[redacted]	10,286	0.10	106,945	194.66
	7.1 No continued effect after 36 months*	[redacted]	[redacted]	[redacted]	[redacted]	11,387	0.16	71,732	97.64
	7.2 OS curves converge over 18 month period (months 36 to 54)*	[redacted]	[redacted]	[redacted]	[redacted]	10,702	0.12	90,824	150.25

* PD utility of 0.60 and mean resource use from KOL survey

6.5 Subgroups

The alternative ERG base case and alternative OS scenario analyses reported in Section 6.4 are repeated for each of the subgroups. As described previously these analyses include the ERG's model adjustments from Section 6.2 and the ERG preferences from section 6.4. The ERG's subgroup analyses were undertaken using Weibull distributions to model PFS. This was different from the manufacturer's subgroup analyses for which they use the Kaplan-Meier curves. However, in the manufacturer's overall base case analysis they justify the use of the Weibull function to smooth the PFS curves. For consistency with the overall population the ERG used the Weibull function to smooth PFS. The difference to the ICER is less than £1000.

1. Weibull distributions were used to model PFS;
2. Acquisition and administration costs are related to the proportion of patients in the stable, on second-line treatment, state at each cycle;
3. The duplicated cycle length adjustment for AEs duration was corrected;
4. Applied a utility decrement over time;
5. The population age was assumed to be 70 years old to estimate the magnitude of age dependent utility decrements;
6. Utility of PD state of 0.60;
7. An additional administration cost for aflibercept of £15;
8. Mean resource use from the KOL survey.

6.5.1 Liver metastases only subgroup

The ICER results of the alternative ERG base case and alternative OS scenario analyses for the liver metastases only subgroup range from £42,895 to £58,257 per QALY (Table 6.9). These results are consistently lower than ICERs for the overall population equivalent scenarios. However, results of the liver metastases only subgroup should be interpreted with caution. Given that OS and PFS independent parametric curves for each treatment were fitted to the respective Kaplan-Meier curves for this subgroup from the VELOUR trial, and that the subgroup corresponded to approximately 25% of the trial population, the analyses referring to this subgroup may be affected by considerable loss of statistical power. This means that is considerable uncertainty surrounding results for this subgroup compared to those of the overall population.

6.5.2 Excluding adjuvant chemotherapy subgroup

The ICER results of the alternative ERG base case and alternative OS scenario analyses for the excluding adjuvant chemotherapy therapy subgroup range from £51,249 to £80,187 per QALY (Table 6.10). These results are consistently lower than ICERs for the overall population equivalent scenarios. Some caution should also be exerted when interpreting these results, as they refer to a *post-hoc* subgroup and may be biased. Furthermore, in the opinion of the clinical advisor to the ERG those patients excluded from this subgroup, i.e. those who underwent adjuvant chemotherapy and relapsed quickly on or after treatment, would not be treated differently to the rest of the trial population in the UK clinical practice.

Table 6.9 Cost effectiveness results for the ERG revised scenarios for the Liver Metastases Only Subgroup

	Scenario	PD utility	Aflibercept + FOLFIRI		FOLFIRI		Inc. Costs (£)	Inc. QALYs	ICER (£/QALY)	% Change
			Mean costs (£)	Mean QALYs	Mean costs (£)	Mean QALYs				
Base case	Liver Metastases Only Subgroup	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	10,974	0.3601	30,474	-
Treatment effect duration	6.1 No continued effect after 30months	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	12,840	0.2907	44,168	44.94
		0.600	[redacted]	[redacted]	[redacted]	[redacted]	12,840	0.2688	47,764	56.74
	6.2 OS curves converge over 18 month period (months 30 to 48)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	11,908	0.2273	52,383	71.89
		0.600	[redacted]	[redacted]	[redacted]	[redacted]	11,908	0.2159	55,159	81.00
	6. OS curves converge over 12 month period (months 30 to 42)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	11,635	0.2081	55,909	83.46
		0.600	[redacted]	[redacted]	[redacted]	[redacted]	11,635	0.1997	58,257	91.17
	7.1 No continued effect after 36 months	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	13,029	0.3037	42,895	40.76
		0.600	[redacted]	[redacted]	[redacted]	[redacted]	13,029	0.2797	46,576	52.84
	7.2 OS curves converge over 18 month period (months 36 to 54)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	12,111	0.2416	50,120	64.47
		0.600	[redacted]	[redacted]	[redacted]	[redacted]	12,111	0.2279	53,136	74.37
	7.3 OS curves converge over 12 month period (months 36 to 48)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	11,909	0.2274	52,368	71.84
		0.600	[redacted]	[redacted]	[redacted]	[redacted]	11,909	0.2160	55,145	80.96

Table 6.10 Cost effectiveness results for the ERG revised scenarios for the Excluding Adjuvant Chemotherapy Subgroup

	Scenario	PD utility	Aflibercept + FOLFIRI		FOLFIRI		Inc. Costs (£)	Inc. QALYs	ICER (£/QALY)	% Change
			Mean costs (£)	Mean QALYs	Mean costs (£)	Mean QALYs				
Base case	ITT Excluding Adjuvant Subgroup	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	8,573	0.2639	32,480	-
Treatment effect duration	6.1 No continued effect after 30months	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	10,084	0.1865	54,072	66.48
		0.600	[redacted]	[redacted]	[redacted]	[redacted]	10,084	0.1678	60,076	84.96
	6.2 OS curves converge over 18 month period (months 30 to 48)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	9,367	0.1379	67,943	109.18
		0.600	[redacted]	[redacted]	[redacted]	[redacted]	9,367	0.1273	73,607	126.62
	6. OS curves converge over 12 month period (months 30 to 42)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	9,144	0.1221	74,863	130.49
		0.600	[redacted]	[redacted]	[redacted]	[redacted]	9,144	0.1140	80,187	146.88
	7.1 No continued effect after 36 months	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	10,292	0.2008	51,249	57.79
		0.600	[redacted]	[redacted]	[redacted]	[redacted]	10,292	0.1799	57,224	76.18
	7.2 OS curves converge over 18 month period (months 36 to 54)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	9,538	0.1499	63,633	95.91
		0.600	[redacted]	[redacted]	[redacted]	[redacted]	9,538	0.1374	69,431	113.77
	7.3 OS curves converge over 12 month period (months 36 to 48)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	9,367	0.1379	67,935	109.16
		0.600	[redacted]	[redacted]	[redacted]	[redacted]	9,367	0.1273	73,599	126.60

6.6 Conclusions

The ERG calculated an alternative base case that included model corrections, updated model programming and the ERG preferred parameters. The ERG corrected the calculations in the model for utility decrements of AEs and updated the model to account for declining HRQoL due to age with the model starting at 70 years old and in accordance to what the ERG believes to be more appropriate assumptions on acquisition and administration costs of second-line treatment. The ERG alternative base case also included a lower estimate of the utility for progressive disease (0.60), a small additional cost to the administration of aflibercept (£15) and used the mean resource use from the KOL survey. These changes increased the ICER to £54,368 per QALY.

The ERG tested a number of alternative OS assumptions. In each of the ERG analyses the manufacturer's base case treatment effect was assumed to hold until either 30 or 36 months. These time points were identified as being highly uncertain in the manufacturer's piecewise model. The ERG explored two assumptions other than the manufacturer's base case as they had only assumed that the treatment effect of aflibercept increased over time. Firstly the ERG tested an assumption where the treatment effect stopped at 30 or 36 months after which the hazard functions were the same between arms. Secondly the ERG tested an assumption where the treatment effect stopped at 30 or 36 months after which the risk of death in the aflibercept + FOLFIRI arm was higher than the FOLFIRI alone arm until the two arms were the same. The second assumption was undertaken to correspond with the increased progression found in the aflibercept + FOLFIRI arm resulting in the PFS curves converging at about 12 months. The ERG's exploration of the OS assumptions resulted in ICERs ranging from £62,894 to £92,089 per QALY. PSA was also conducted on selected OS scenarios. ICERs calculated from the PSA were found to be higher than the deterministic ICERs of the same scenarios.

The ERG did further analyses in the subgroups presented by the manufacturer. Combining the ERG's OS assumptions and the ERG's alternative base case (defined in Section 6.5) the ICER of aflibercept + FOLFIRI vs. FOLFIRI ranged from £46,576 to £58,257 per QALY in the liver metastases only subgroup and from £57,224 to £80,187 per QALY in the excluding adjuvant therapy subgroup.

7 END OF LIFE

The average life expectancy of patients with mCRC is less than 24 months. The MS states that the 5-year survival rate for patients receiving second-line mCRC therapy is estimated at 10%, and observed median OS ranges between 10.5 and 18 months; the MS references Seymour *et al.*, 2011,[2] Jonker *et al.*, 2007,[3] Peeters *et al.*, 2010[4] and Kozloff *et al.*, 2011.[5] The study by Seymour *et al.*[2] was conducted in the UK, so is likely to be the most reflective of the UK population; the median OS for this trial was 10.5 months.

The VELOUR trial found a small, but statistically significant, increase in overall survival (OS) for patients in the aflibercept group compared with patients in the placebo group; the difference in median OS was 1.44 months (13.5 months in the aflibercept group and 12.06 months in the placebo group). The MS included an estimation of mean OS benefit based on an extrapolation of the data: 4.7 months. This was considerably longer than the median OS benefit reported (1.44 months).The ERG has expressed concerns on the robustness of the evidence regarding aflibercept's survival gains, especially for gains accrued in the extrapolated portion of the OS curves. As discussed previously, the manufacturer's base case assumptions on OS imply the maintenance of a differential treatment effect throughout the time horizon. In Section 6.3.1, the ERG demonstrated the model's sensitivity to alternative (more conservative) assumptions regarding the magnitude and duration of the treatment effect on OS. The mean OS gain for aflibercept in the ERG alternative scenarios, which assumed no continued treatment effect for both treatment groups after 30 or 36 months ranged from 3.4 to 3.7 months. Scenarios assuming that the OS curves converged had a mean OS gain for aflibercept ranging from 2.1 to 2.6 months.

The number of patients indicated for treatment with aflibercept alongside FOLFIRI in the UK was estimated by the manufacturer to be small. The MS states that an estimated 4,028 patients will receive second-line treatment for mCRC in England and Wales in 2013. Furthermore, the MS states that around 64% of patients receive oxaliplatin as first-line therapy; therefore the manufacturer estimates that approximately 2,578 mCRC patients would be eligible for aflibercept as second-line therapy, as per the licenced indication.

Despite the argument in the MS that with the inclusion of the PAS and in the end of life setting "the ICER for aflibercept is £36,294, indicating aflibercept can be considered a cost-effective treatment", the ERG additional analyses in Section 6 demonstrate that this estimate is very uncertain. The ICERs of the alternative ERG base case combined with the alternative OS scenarios described in Section 6.4 range between £47,965 and £92,089 per QALY for the overall population.

8 OVERALL CONCLUSIONS

Evidence from one good quality RCT demonstrates a statistically significant, but clinically small (difference in median OS 1.44 months), increase in OS and PFS with aflibercept plus FOLFIRI, compared with FOLFIRI alone. However, this level of benefit may not be achieved in clinical practice, as the trial participants were potentially younger and healthier than the population eligible for aflibercept plus FOLFIRI in UK practice. Grade 3-4 adverse events are common with aflibercept plus FOLFIRI (83.5% compared with 62.5% in the placebo group), leading to discontinuation of treatment in a large number of patients (26.8% in the aflibercept group compared with 12.1% in the placebo group). There are currently no data on HRQoL for patients taking aflibercept.

No previously published cost-effectives results were pertinent to the decision problem. The manufacturer's *de novo* cost-effectiveness analysis met the criteria for the NICE reference case. After correcting the model programming and updating the model to include the ERG's preferred parameter estimates the ICER from the ERG's alternative base case was £54,368 per QALY. Further analyses were conducted to explore the uncertainty around the extrapolation of OS. All alternative OS scenarios tested by the ERG increased the ICER further. Scenario analyses conducted on the manufacturer's subgroups resulted in ICERs lower than the ITT population, with some scenarios in the liver metastases only subgroup being less than £50,000 per QALY. The ERG was provided with insufficient information on the manufacturer's PAS to verify that the PAS was implemented correctly in the model.

8.1 Implications for research

Future studies of aflibercept should assess possible variations in effect and predictors of treatment efficacy, such as biomarkers. This would allow better selection of patients more likely to benefit from aflibercept.

References

1. Hoyle M, Crathorne L, Peters J, Jones-Hughes T, Cooper C, Napier M, et al. *The 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 150 and part review of technology appraisal 118): a systematic review and economic model. NIHR HTA Programme*. Exeter: Peninsula Technology Assessment Group (PenTAG), University of Exeter, ; 2011.
2. Seymour MT, Brown SR, Richman S, Middleton GW, Maughan T, Olivier C, et al. Addition of panitumumab to irinotecan: Results of PICCOLO, a randomized controlled trial in advanced colorectal cancer (aCRC). *Journal of Clinical Oncology. Conference: ASCO Annual Meeting* 2011;**29**. Available from: http://meeting.ascopubs.org/cgi/content/abstract/29/15-suppl/3523?sid=40db32b0-7c00-46ac-8e7c-12a198b1f56e
3. Jonker DJ, O'Callaghan CJ, Karapetis CS, Zalcberg JR, Tu D, Au HJ, et al. Cetuximab for the treatment of colorectal cancer. *N Engl J Med* 2007;**357**:2040-8.
4. Peeters M, Price TJ, Cervantes A, Sobrero AF, Ducreux M, Hotko Y, et al. Randomized phase III study of panitumumab with fluorouracil, leucovorin, and irinotecan (FOLFIRI) compared with FOLFIRI alone as second-line treatment in patients with metastatic colorectal cancer. *J Clin Oncol* 2010;**28 (31)**:4706-13.
5. Kozloff M, Bekaii-Saab TS, Bendell JC, Cohn AL, Hurwitz H, Roach N, et al. Effectiveness of first- or second-line bevacizumab treatment in elderly patients with metastatic colorectal cancer in ARIES, an observational cohort study [abstract 3625]. *J Clin Oncol.* 2011;**29 (Suppl)**.
6. National Institute for Health and Care Excellence (NICE). *Colorectal cancer: The diagnosis and management of colorectal cancer. CG131*. [s.l.]; 2011. Available from: http://www.nice.org.uk/nicemedia/live/13597/56998/56998.pdf
7. Tournigand C, Andre T, Achille E, Lledo G, Flesh M, Mery-Mignard D, et al. FOLFIRI followed by FOLFOX6 or the reverse sequence in advanced colorectal cancer: a randomized GERCOR study. *J Clin Oncol* 2004;**22**:229-37.
8. National Institute for Health and Care Excellence (NICE). *Cetuximab, bevacizumab and panitumumab for the treatment of metastatic colorectal cancer after first-line chemotherapy. TA242*. [s.l.]; 2012. Available from: http://www.nice.org.uk/nicemedia/live/13651/57924/57924.pdf
9. Adams RA, Meade AM, Seymour MT, Wilson RH, Madi A, Fisher D, et al. Intermittent versus continuous oxaliplatin and fluoropyrimidine combination chemotherapy for first-line treatment of advanced colorectal cancer: results of the randomised phase 3 MRC COIN trial. *Lancet Oncol* 2011;**12**:642-53.
10. National Institute for Health and Clinical Excellence (NICE). *Cetuximab, bevacizumab and panitumumab for the treatment of metastatic colorectal cancer after first-line chemotherapy. TA242*. London: NICE; 2012.
11. Hurwitz HI, Yi J, Ince W, Novotny WF, Rosen O. The clinical benefit of bevacizumab in metastatic colorectal cancer is independent of K-ras mutation status: analysis of a phase III study of bevacizumab with chemotherapy in previously untreated metastatic colorectal cancer. *Oncologist* 2009;**14**:22-8.
12. Ince WL, Jubb AM, Holden SN, Holmgren EB, Tobin P, Sridhar M, et al. Association of k-ras, b-raf, and p53 status with the treatment effect of bevacizumab. *J Natl Cancer Inst* 2005;**97**:981-9.
13. Bennouna J, Sastre J, Arnold D, Osterlund P, Greil R, Van Cutsem E, et al. Continuation of bevacizumab after first progression in metastatic colorectal cancer (ML18147): a randomised phase 3 trial. *Lancet Oncol* 2013;**14**:29-37.

14. European Medicines Agency (EMA). *Avastin : EPAR. Annex I: Summary of product characteristics*; 2013. Available from: http://www.emea.europa.eu/docs/en_GB/document_library/EPAR_-_Product_Information/human/000582/WC500029271.pdf
15. Giantonio BJ, Catalano PJ, Meropol NJ, O'Dwyer PJ, Mitchell EP, Alberts SR, et al. Bevacizumab in combination with oxaliplatin, fluorouracil, and leucovorin (FOLFOX4) for previously treated metastatic colorectal cancer: results from the Eastern Cooperative Oncology Group Study E3200. *J Clin Oncol* 2007;**25**:1539-44.
16. Beretta GD, Petrelli F, Stinco S, Cabiddu M, Ghilardi M, Squadroni M, et al. FOLFIRI + bevacizumab as second-line therapy for metastatic colorectal cancer pretreated with oxaliplatin: a pooled analysis of published trials. *Med Oncol* 2013;**30**:486.
17. Peeters M, Strickland AH, Lichinitser M, Suresh AVS, Manikhas G, Shapiro J, et al. A randomised, double-blind, placebo-controlled phase 2 study of trebananib (AMG 386) in combination with FOLFIRI in patients with previously treated metastatic colorectal carcinoma. *Br J Cancer* 2013;**108**:503-11.
18. Thomas AL, Wilson RH, Steward WP, Adams RA, Bridges SE, Casbard AC, et al. A randomized phase II study of irinotecan, 5-fluorouracil, and folinic acid (FOLFIRI) with or without the addition of an endothelin receptor antagonist in patients with metastatic colorectal cancer after failure of oxaliplatin-containing chemotherapy. *Journal of Clinical Oncology. Conference: ASCO Annual Meeting* 2011;**29**.
19. Davis S, Tappenden P, A. C. *A review of studies examining the relationship between progression-free survival and overall survival in advanced or metastatic cancer: Report by the Decision Support Unit*: Decision Support Unit, ScHARR,
University of Sheffield; 2012. Available from: http://www.nicedsu.org.uk/PFSOS%20Report.FINAL.06.08.12.pdf
20. Briggs A, Claxton K, Sculpher MJ, Andrew B. *Decision modelling for health economic evaluation*. Oxford: Oxford University Press; 2006.
21. European Medicines Agency (EMA). *Appendix 1 to the guideline on the evaluation of anticancer medicinal products in man; Methodological consideration for using progression-free survival (PFS) or disease-free survival (DFS) in confirmatory trials.* . London; 2011. Available from: http://www.emea.europa.eu/docs/en_GB/document_library/Scientific_guideline/2013/01/WC500137126.pdf
22. Soria JC, Massard C, Le Chevalier T. Should progression-free survival be the primary measure of efficacy for advanced NSCLC therapy? *Ann Oncol* 2010;**21**:2324-32.
23. U. S. Food and Drug Administration (FDA). *Guidance for industry clinical trial endpoints for the approval of cancer drugs and biologics*; 2007.
24. Sundar S, Wu J, Hillaby K, Yap J, Lilford R. A systematic review evaluating the relationship between progression free survival and post progression survival in advanced ovarian cancer. *Gynecol Oncol* 2012;**125**:493-9.
25. Chirila C, Odom D, Devercelli G, Khan S, Sherif BN, Kaye JA, et al. Meta-analysis of the association between progression-free survival and overall survival in metastatic colorectal cancer. *Int J Colorectal Dis* 2012;**27**:623-34.
26. Giessen C, Laubender RP, Ankerst DP, Stintzing S, Modest DP, Mansmann U, et al. Progression-free survival as a surrogate endpoint for median overall survival in metastatic colorectal cancer: literature-based analysis from 50 randomized first-line trials. *Clin Cancer Res* 2013;**19**:225-35.
27. Petrelli F, Barni S. Correlation of progression-free and post-progression survival with overall survival in advanced colorectal cancer. *Ann Oncol* 2013;**24**:186-92.
28. Cancer Research UK. *Bowel cancer incidence statistics*. 2011. [cited 2013 20-03]. Available from: http://www.cancerresearchuk.org/cancer-info/cancerstats/types/bowel/incidence/.
29. Sobrero AF, Maurel J, Fehrenbacher L, Scheithauer W, Abubakr YA, Lutz MP, et al. EPIC: Phase III Trial of Cetuximab Plus Irinotecan After Fluoropyrimidine and Oxaliplatin Failure in Patients

With Metastatic Colorectal Cancer. *J Clin Oncol* 2008;**26**:2311-9. Available from: http://jco.ascopubs.org/content/26/14/2311.abstract
30. Higgins JP, Altman DG, Gotzsche PC, Juni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 2011;**343**:d5928.
31. U.S. Food and Drug Administration (FDA). *ZALTRAP (ziv-aflibercept): Statistical Review*; 2012. Available from: http://www.accessdata.fda.gov/drugsatfda_docs/nda/2012/125418Orig1s000StatR.pdf
32. European Medicines Agency (EMA). *Zaltrap : EPAR - Public assessment report* 2013. Available from: http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Public_assessment_report/human/002532/WC500139710.pdf
33. U.S. Food and Drug Administration (FDA). *ZALTRAP (ziv-aflibercept): Medical Review*; 2012. Available from: http://www.accessdata.fda.gov/drugsatfda_docs/nda/2012/125418Orig1s000MedR.pdf
34. Sanofi Aventis. A Multinational, Randomized, Double-blind Study, Comparing the Efficacy of Aflibercept Once Every 2 Weeks versus Placebo in Patients with Metastatic Colorectal Cancer (MCRC) Treated with Irinotecan / 5-FU Combination (FOLFIRI) after Failure of an oxaliplatin based regimen. STUDY NUMBER: EFC10262. 2011.
35. Morlock R, Yu E, Ray J. A cost-effectiveness analysis of bevacizumab (BV) plus chemotherapy (CT) versus aflibercept (AFLI) plus CT in patients with metastatic colorectal cancer (mCRC) previously treated with BV. In: *J Clin Oncol (Meeting Abstracts)*; 2013. 2013. p. 417.
36. Philips Z, Ginnelly L, Sculpher M, Claxton K, Golder S, Riemsma R, et al. *Review of guidelines for good practice in decision-analytic modelling in health technology assessment*: York Pub. Services; 2004.
37. Dolan P. Modeling valuations for EuroQol health states. *Med Care* 1997:1095-108.
38. Joint Formulary Committee. *British National Formulary (BNF) [Internet]*. 2011. [cited Available from: http://www.bnf.org/bnf/go?bnf/current/.
39. Latimer N. NICE DSU technical support document 14: survival analysis for economic evaluations alongside clinical trials-extrapolation with patient-level data. *Sheffield: Report by the Decision Support Unit* 2011.
40. National Institute for Clinical and Clinical Excellence (NICE). *Guide to the methods of technology appraisal*. London: NICE; 2008. Available from: http://www.nice.org.uk/media/B52/A7/TAMethodsGuideUpdatedJune2008.pdf
41. Bennett L, Zhao Z, Barber B, Zhou X, Peeters M, Zhang J, et al. Health-related quality of life in patients with metastatic colorectal cancer treated with panitumumab in first- or second-line treatment. *Br J Cancer* 2011;**105**:1495-502.
42. National Institute for Health and Clinical Excellence (NICE). *Bevacizumab in combination with oxaliplatin and either fluorouracil plus folinic acid or capecitabine for the treatment of metastatic colorectal cancer. TA 212*. London: NICE; 2010. Available from: http://www.nice.org.uk/nicemedia/live/13291/52091/52091.pdf
43. Whyte S, Pandor A, Stevenson M, Rees A. *Bevacizumab in combination with Fluoropyrimidine-based chemotherapy for the first-line treatment of metastatic colorectal cancer – a single technology appraisal. NIHR HTA Programme. TA212*. Sheffield: School of Health and Related Research (ScHARR), The University of Sheffield; 2009.
44. Odom D, Barber B, Bennett L, Peeters M, Zhao Z, Kaye J, et al. Health-related quality of life and colorectal cancer-specific symptoms in patients with chemotherapy-refractory metastatic disease treated with panitumumab. *Int J Colorectal Dis* 2011;**26**:173-81.
45. Best JH, Garrison LP, Hollingworth W, Ramsey SD, Veenstra DL. Preference values associated with stage III colon cancer and adjuvant chemotherapy. *Qual Life Res* 2010;**19**:391-400.
46. National Collaborating Centre for Cancer (NCC-C). *Colorectal cancer: the diagnosis and management of colorectal cancer. Evidence Review*. [s.l.]: NCC-C; 2011.

47. Sullivan PW, Slejko JF, Sculpher MJ, Ghushchyan V. Catalogue of EQ-5D scores for the United Kingdom. *Med Decis Making* 2011;**31**:800-4 %@ 0272-989X.
48. Kind P, Hardman G, Macran S. *UK population norms for EQ-5D*: Centre for Health Economics, University of York; 1999.
49. 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. Available from: http://www.shef.ac.uk/polopoly_fs/1.43391!/file/HEDS-DP-09-12.pdf
50. Tolley K, Goad C, Yi Y, Maroudas P, Haiderali A, Thompson G. Utility elicitation study in the UK general public for late-stage chronic lymphocytic leukaemia. *Eur J Health Econ* 2012:1-11 %@ 1618-7598.
51. Shabaruddin FH, Elliott RA, Valle JW, Newman WG, Payne K. Understanding chemotherapy treatment pathways of advanced colorectal cancer patients to inform an economic evaluation in the United Kingdom. *Br J Cancer* 2010;**103**:315-23. Available from: http://dx.doi.org/10.1038/sj.bjc.6605766
52. UK National Health System (NHS). *NHS reference costs 2011-12. National schedules of Reference Costs: NHS Own Costs*. Department of Health (DH); 2012. [cited 12-04-2013]. Available from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/127115/NSRC01-2011-12.xls.xls
53. UK National Health System (NHS). *NHS reference costs 2010-11. NHS Trusts and PCTs combined reference cost schedules 2010-11*. Department of Health (DH); 2011. [cited 12-04-2013]. Available from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/146974/dh_131148.xls.xls
54. Curtis L. *Unit Costs of Health and Social Care 2012*. Kent: Personal Social Services Research Unit (PPSRU),The University of Kent; 2012. Available from: http://www.pssru.ac.uk/project-pages/unit-costs/2012/index.php?file=full
55. UK National Health System (NHS). *Confirmation of Payment by Results (PbR) arrangements for 2011-12*. Department of Health (DH); 2011. [cited 12-04-2013]. Available from: https://www.gov.uk/government/publications/confirmation-of-payment-by-results-pbr-arrangements-for-2011-12.
56. Joint Formulary Committee. *Joint Formulary Committee. British National Formulary (online).* BMJ Group and Pharmaceutical Press; 2013. [cited 12-04-2013]. Available from: http://www.medicinescomplete.com.
57. National Institute for Health and Clinical Excellence (NICE). *Bevacizumab and cetuximab for the treatment of metastatic colorectal cancer. TA118*. [s.l]: NICE; 2007. Available from: http://www.nice.org.uk/TA118
58. Tappenden P, Jones R, Paisley S, Carroll C. *The use of bevacizumab and cetuximab for the treatment of metastatic colorectal cancer. TA118. NIHR HTA Programme*. Sheffield: The University of Sheffield; 2006.

Appendix 1 Cost effectiveness results for the ERG alternative scenarios assuming median (KOL survey) resource use

	Scenario	PD utility	Aflibercept + FOLFIRI		FOLFIRI		Inc. Costs (£)	Inc. QALYs	ICER (£/QALY)	% Change
			Mean costs (£)	Mean QALYs	Mean costs (£)	Mean QALYs				
Base case	Not applicable	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	8,816	0.2429	36,294	-
Treatment effect duration	6.1 No continued effect after 30months	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	8,984	0.1645	54,607	50.46
		0.600	[redacted]	[redacted]	[redacted]	[redacted]	8,984	0.1478	60,788	67.49
	6.2 OS curves converge over 18 month period (months 30 to 48)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	8,603	0.1188	72,412	99.52
		0.600	[redacted]	[redacted]	[redacted]	[redacted]	8,603	0.1096	78,470	116.21
	6. OS curves converge over 12 month period (months 30 to 42)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	8,484	0.1039	81,630	124.91
		0.600	[redacted]	[redacted]	[redacted]	[redacted]	8,484	0.0971	87,353	140.68
	7.1 No continued effect after 36 months	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	9,099	0.1785	50,968	40.43
		0.600	[redacted]	[redacted]	[redacted]	[redacted]	9,099	0.1595	57,042	57.17
	7.2 OS curves converge over 18 month period (months 36 to 54)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	8,693	0.1300	66,859	84.22
		0.600	[redacted]	[redacted]	[redacted]	[redacted]	8,693	0.1191	73,011	101.17
	7.3 OS curves converge over 12 month	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	8,602	0.1186	72,519	99.81

	Scenario	**PD utility**	**Aflibercept + FOLFIRI**		**FOLFIRI**		**Inc. Costs (£)**	**Inc. QALYs**	**ICER (£/QALY)**	**% Change**
			Mean costs (£)	**Mean QALYs**	**Mean costs (£)**	**Mean QALYs**				
Base case	Not applicable	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	8,816	0.2429	36,294	-
	period (months 36 to 48)	0.600	[redacted]	[redacted]	[redacted]	[redacted]	8,602	0.1095	78,575	116.50

Appendix 2 Quality assessment using the Philips[36] economic modelling checklist

Quality criterion	Question(s)	Response (√, X, or NA)	Comments
S1	Is there a clear statement of the decision problem?	√	
	Is the objective of the evaluation and model specified and consistent with the stated decision problem?	√	
	Is the primary decision-maker specified?	√	NHS and Personal Social Services
S2	Is the perspective of the model stated clearly?	√	
	Are the model inputs consistent with the stated perspective?	√	
	Has the scope of the model been stated and justified?	√	
	Are the outcomes of the model consistent with the perspective, scope and overall objective of the model?	√	
S3	Is the structure of the model consistent with a coherent theory of the health condition under evaluation?	√	
	Are the sources of data used to develop the structure of the model specified?	√	MS states that the model structure is common in economic evaluation of oncology treatments, and is also consistent with those used in previous health technology assessments for the treatment of mCRC.
	Are the causal relationships described by the model structure justified appropriately?	√	
S4	Are the structural assumptions transparent and justified?	√	
	Are the structural assumptions reasonable given the overall objective, perspective and scope of the model?	√	
S5	Is there a clear definition of the options under evaluation?	√	
	Have all feasible and practical options been evaluated?	√	
	Is there justification for the exclusion of feasible options?	NA	
S6	Is the chosen model type appropriate given the decision problem and specified causal relationships within the model?	√	
S7	Is the time horizon of the model sufficient to reflect all important differences between options?	√	
	Are the time horizon of the model, the duration of treatment and the duration of treatment effect described and justified?	X	The economic model follows a time horizon of 15 years. Less than 0.01% of patients are expected to survive beyond this period, using the manufacturer's chosen extrapolation methods. Treatment duration for each treatment arm is assumed to be the same as in the VELOUR trial. Treatment effect is implicitly assumed to last beyond treatment duration and throughout the time horizon, due to the manufacturer's selected OS extrapolation approach. This assumption was not justified by the manufacturer.
S8	Do the disease states (state transition model) or the pathways (decision tree model) reflect the underlying biological process of the disease in question and the impact of interventions?	√	
S9	Is the cycle length defined and justified in terms of the natural history of disease?	X	Two weeks cycle length was justified based on the biweekly treatment schedules, and not

Quality criterion	Question(s)	Response (√, X, or NA)	Comments
			on disease history.
D1	Are the data identification methods transparent and appropriate given the objectives of the model?	√	
	Where choices have been made between data sources, are these justified appropriately?	√	
	Has particular attention been paid to identifying data for the important parameters in the model?	√	
	Has the quality of the data been assessed appropriately?	√	
	Where expert opinion has been used, are the methods described and justified?	X	Median of the KOL survey was used to inform resource use related with community and social and personal care disease management, despite mean being the preferred estimates in cost-effectiveness analysis. This was justified based on the skew of the data. However, the manufacturer did not further explain why the data is distributed in this way and whether the skew may be reflective of resource use in clinical practice. Moreover, the use of the median is inconsistent, as mean estimates for other parameters were estimated from the KOL survey and used to inform resource use for the treatment of AEs.
D2	Is the data modelling methodology based on justifiable statistical and epidemiological techniques?	√	
D2a	Is the choice of baseline data described and justified?	NA	OS and PFS are modelled by fitting independent parametric curves to empirical data from the VELOUR trial. No external baseline data was considered.
	Are transition probabilities calculated appropriately?	√	
	Has a half-cycle correction been applied to both cost and outcome?	X	Half-cycle correction was applied to costs and QALYs, with the exception of acquisition and administration costs of second-line treatment.
	If not, has this omission been justified?	X	This was not explicitly justified, but acquisition and administration costs of second-line treatment were applied to all patients for a fixed number of cycles and were not attached to any particular health state. Therefore, there was no need to account for patients that may have transitioned health state at intermediate time points within the model cycle.
D2b	If relative treatment effects have been derived from trial data, have they been synthesised using appropriate techniques?	NA	There was only one RCT trial identified as relevant to the appraisal.
	Have the methods and assumptions used to extrapolate short-term results to final outcomes been documented and justified?	√	OS data was extrapolated beyond trial data by fitting independent parametric survival functions.
	Have alternative extrapolation assumptions been explored through sensitivity analysis?	X	Different parametric distributions were modelled and tested. However, the ERG believes that there were alternative survival (clinically plausible) extrapolation models that could have been applied to explore uncertainty.
	Have assumptions regarding the continuing effect of treatment once treatment is complete been documented and justified?	X	
D2c	Are the costs incorporated into the model justified?	√	
	Has the source for all costs been described?	√	
	Have discount rates been described and justified given the target decision-maker?	√	
D2d	Are the utilities incorporated into the model appropriate?	X	There are concerns that the population of the mCRC Utilities study used to inform HRQoL estimates in the model may not be reflective of the mCRC patients in UK routine clinical practice. The study population was younger and potentially fitter than the mCRC patients

Quality criterion	Question(s)	Response (√, X, or NA)	Comments
			in UK routine clinical practice. Moreover, the model did not allow for utility decrements associated with ageing, and utility in PD state may not be reflective of the entire post-progression period.
	Is the source for the utility weights referenced?	√	
	Are the methods of derivation for the utility weights justified?	√	
D3	Have all data incorporated into the model been described and referenced in sufficient detail?	X	Unit costs were described, but there were some discrepancies regarding the sources referenced in the MS.
	Has the use of mutually inconsistent data been justified (i.e. are assumptions and choices appropriate)?	NA	
	Is the process of data incorporation transparent?	√	
	If data have been incorporated as distributions, has the choice of distribution for each parameter been described and justified?	X	Probabilistic distributions attached to model parameters were not described for every parameter, and justification for their selection was not provided by the manufacturer.
	If data have been incorporated as distributions, is it clear that second order uncertainty is reflected?	X	Standard errors were calculated based on a seemingly arbitrary 30% variation around the mean, for the following parameters: administration costs per cycle, weighted average costs and disutilities for AEs, dosage per administration for the irinotecan component, proportion of patients on further lines of treatment, and aggregated drug acquisition costs.
D4	Have the four principal types of uncertainty been addressed?	X	Structural uncertainty was not fully characterised.
	If not, has the omission of particular forms of uncertainty been justified?	X	
D4a	Have methodological uncertainties been addressed by running alternative versions of the model with different methodological assumptions?	X	The manufacturer did not thoroughly explore alternative approaches to model OS data extrapolation.
D4b	Is there evidence that structural uncertainties have been addressed via sensitivity analysis?	X	Despite having conducted scenario analyses so as to characterise structural uncertainty, the manufacturer did not consider the joint effect of related assumptions.
D4c	Has heterogeneity been dealt with by running the model separately for different subgroups?	√	
D4d	Are the methods of assessment of parameter uncertainty appropriate?	X	Calculation of standard errors based an assumed 30% variation was not justified (see comments on D3).
	If data are incorporated as point estimates, are the ranges used for sensitivity analysis stated clearly and justified?	X	All parameters in the univariate sensitivity analysis were varied within a [redacted] range in both directions. No justification for the selection of the range was provided.
C1	Is there evidence that the mathematical logic of the model has been tested thoroughly before use?	√	
C2	Are any counterintuitive results from the model explained and justified?	NA	
	If the model has been calibrated against independent data, have any differences been explained and justified?	NA	
	Have the results of the model been compared with those of previous models and any differences in results explained?	NA	No other models were identified in the manufacturer's review of the literature.

AEs, Adverse Events; HRQoL, Health Related Quality of Life; KOL, Key Opinion Leaders; mCRC, metastatic Colorectal Cancer; MS, Manufacturer's Submission, PD, Progressive Disease: RCT – Randomised Clinical Trial