NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

Final appraisal determination

Aflibercept in combination with irinotecan and fluorouracil-based therapy for treating metastatic colorectal cancer that has progressed following prior oxaliplatin-based chemotherapy

This guidance was developed using the single technology appraisal (STA) process.

1 Guidance

1.1 Aflibercept in combination with irinotecan and fluorouracil-based therapy is not recommended within its marketing authorisation for treating metastatic colorectal cancer that is resistant to or has progressed after an oxaliplatin-containing regimen.

1.2 People currently receiving aflibercept in combination with irinotecan and fluorouracil-based therapy for treating metastatic colorectal cancer that is resistant to or has progressed after an oxaliplatin-containing regimen should be able to continue treatment until they and their clinician consider it appropriate to stop.

2 The technology

2.1 Aflibercept (Zaltrap, Sanofi) is a recombinant human fusion protein that blocks the vascular endothelial growth factor (VEGF) pathway by preferentially binding to VEGF-A, VEGF-B and placental growth factor, which play an important role in the formation of new blood vessels in solid tumours (angiogenesis). By preventing these

factors from activating their endogenous receptors, aflibercept interferes with the process by which blood vessels and capillaries expand into tumours (vascularisation), and so inhibits tumour growth. Aflibercept in combination with folinic acid/5-fluorouracil/irinotecan (FOLFIRI) (that is, in combination with irinotecan and fluorouracil-based therapy) has a UK marketing authorisation 'for the treatment of adults with metastatic colorectal cancer that is resistant to or has progressed after an oxaliplatin-containing regimen'. The summary of product characteristics states that aflibercept should be administered as an intravenous infusion over 1 hour at a dose of 4 mg/kg of body weight, followed by the FOLFIRI regimen, every 2 weeks until the disease progresses or unacceptable toxicity occurs.

2.2 The summary of product characteristics lists the following most common adverse reactions (according to the Common Terminology Criteria for Adverse Events v3.0) for aflibercept plus FOLFIRI in order of decreasing frequency: leukopenia, diarrhoea, neutropenia, proteinuria, increased plasma activity of aspartate aminotransferase, stomatitis, fatigue, thrombocytopenia, increased plasma activity of alanine aminotransferase, hypertension, weight loss, decreased appetite, epistaxis, abdominal pain, dysphonia, increased serum creatinine and headache. For full details of adverse reactions and contraindications, see the summary of product characteristics.

2.3 The manufacturer states that the net price of a vial of 100 mg aflibercept is £295.65, and the net price of a vial of 200 mg aflibercept is £591.30. The cost per patient will vary with dose adjustment and treatment duration. The manufacturer of aflibercept (Sanofi) has agreed a patient access scheme with the Department of Health that makes aflibercept available with a discount. The size

of the discount is commercial in confidence. The Department of Health considered that this patient access scheme does not constitute an excessive administrative burden on the NHS. Costs may vary in different settings because of negotiated procurement discounts.

3 The manufacturer's submission

The Appraisal Committee (section 9) considered evidence submitted by the manufacturer of aflibercept and a review of this submission by the Evidence Review Group (ERG; section 10).

Clinical-effectiveness evidence

3.1 The manufacturer did a systematic literature review of studies evaluating the efficacy and safety of second-line treatments for metastatic colorectal cancer. It identified 1 relevant randomised controlled trial (RCT), the VELOUR trial, from which it obtained the key clinical evidence. The VELOUR trial was a double-blind placebo-controlled phase III study that was conducted in 176 centres in 28 countries, including the UK. Eligible patients were adults who had inoperable metastatic colorectal cancer, and whose disease progressed on or after treatment with only 1 prior oxaliplatin-based chemotherapy regimen. Investigators randomised patients in a 1:1 ratio to either aflibercept plus folinic acid/5-fluorouracil/irinotecan (FOLFIRI) (n=612) or placebo plus FOLFIRI (n=614). They stratified randomisation by patients' wellbeing and ability to perform daily activities using the Eastern Cooperative Oncology Group Performance Status (ECOG PS), and whether or not the patient had received prior therapy with bevacizumab. Patients received either aflibercept at a dose of 4 mg/kg or placebo over 1 hour on day 1, every 2 weeks, both intravenously, immediately followed by FOLFIRI. During the trial,

patients could stop 1 study treatment (aflibercept or placebo, or FOLFIRI) but still receive the other components of the regimen. Treatment continued until disease progressed, unacceptable toxicity occurred, or the patient declined further treatment.

3.2 The primary end point in the VELOUR trial was overall survival, defined as time from randomisation to death from any cause. One of the secondary end points was progression-free survival as assessed by an independent review committee based on radiologic progression; it was determined as time from randomisation to first observation of disease progression (at least a 20% increase in the sum of the longest diameter of target tumours, the unequivocal increase in the size of non-target tumours or the appearance of 1 or more new tumours), or death from any cause. In addition, disease progression determined by local investigators was recorded during the trial. Other secondary end points were objective response (complete and partial responses) according to Response Evaluation Criteria In Solid Tumors criteria version 1, and adverse events and abnormal laboratory findings.

3.3 The manufacturer stated that patient characteristics and disease history at baseline were well balanced between the aflibercept and placebo groups. Of the patients randomised in the study, the median age was 61 years, 58.6% were men, 97.8% had a baseline ECOG PS of 0 or 1, and 2.2% had a baseline ECOG PS of 2. The marketing authorisation for aflibercept stipulates prior treatment with an oxaliplatin-containing regimen. In the VELOUR trial, 90.2% of patients randomised to aflibercept plus FOLFIRI and 89.4% of those randomised to placebo plus FOLFIRI had received prior oxaliplatin-based chemotherapy for locally advanced or metastatic disease. Approximately 10% of patients had received prior oxaliplatin-based chemotherapy in the adjuvant setting (that is, as

an additional treatment given after the primary treatment). Oxaliplatin-based regimens were given in combination with bevacizumab in 30.4% of patients.

3.4 The manufacturer determined that it needed 863 death events to detect a statistically significant 20% risk reduction in the aflibercept group compared with the placebo group; this determined the study cut-off date. To estimate time-to-event parameters (overall survival and progression-free survival), the manufacturer used survival analysis. It calculated hazard ratios and confidence intervals for the primary and subgroup analyses using a Cox proportional hazards model. It also established heterogeneity of treatment effect among subgroups using a Cox proportional hazards model, and provided an interaction test for each subgroup analysis. If a patient neither died nor had disease progression during the trial, the manufacturer censored the patient at the date when the tumour was last assessed or at the study cut-off date.

3.5 The median follow-up for the overall population at the time of the primary analysis was 22.28 months, with the longest follow-up being 36 months. At the study cut-off date, 403 patients (65.8%) randomised to aflibercept and 460 patients (74.9%) randomised to placebo had died. Median overall survival was estimated to be 1.44 months longer for aflibercept than placebo (aflibercept 13.50 months, placebo 12.06 months), and the corresponding hazard ratio was 0.817 (95.34% confidence interval [CI] 0.713 to 0.937, p=0.0032), suggesting a reduction in the risk of death of 18.3% with aflibercept compared with placebo. The probabilities of overall survival at 6, 12, 18, 24 and 30 months were consistently higher in the aflibercept group than in the placebo group; the probability of overall survival was 4% higher at 6 months, and 85% higher at 30 months.

3.6 The manufacturer noted that the Kaplan–Meier curves for overall survival separated early and continued to separate over time, and suggested that there were patients who experienced a sustained benefit after treatment with aflibercept. Because of this, the manufacturer indicated that the difference in median overall survival of 1.44 months may underestimate the overall clinical benefit of adding aflibercept to FOLFIRI. In addition, the manufacturer calculated hazard ratios for overall survival by 6-month periods up to 18 months after randomisation, and it combined all time points thereafter into a single hazard ratio. This analysis showed that hazard ratios improved over time, implying that the difference in overall survival increased in favour of aflibercept the longer patients received treatment. In response to a clarification request by the ERG, the manufacturer provided hazard ratios and the number of patients at risk of dying 18 months after randomisation by 6-month periods. These hazard ratios continued to decrease over time (suggesting that the difference in overall survival continued to increase in favour of aflibercept), but had confidence intervals that crossed 1.00 (that is, the differences were not statistically significant).

3.7 The manufacturer estimated the mean overall survival by fitting separate parametric functions to the trial data for each treatment group, and extrapolating to provide complete curves (given that calculating the mean required all patients to have died). It modelled each treatment group separately, rather than modelling treatment as a covariate, because the log-cumulative hazard plots (used to evaluate the assumption that a hazard ratio between 2 treatments remains constant over time) were not parallel and crossed. The manufacturer considered that the log-logistic function provided the best fit for overall survival for both treatment groups. The

log-logistic function, however, gave a long tail (implying that some patients would live implausibly long), so the manufacturer truncated the curves at 15 years after randomisation (this assumed that all patients die by 15 years). Using this approach, the manufacturer estimated that aflibercept would extend mean overall survival by 4.7 months compared with placebo (aflibercept 22.8 months, placebo 18.1 months); without truncating the survival curves, the difference in mean overall survival was 6.6 months. In response to a clarification request by the ERG, the manufacturer provided estimates with the analysis truncated at 5 and 10 years. The manufacturer designated the results of this analysis as academic in confidence. The manufacturer also provided 'restricted' mean overall survivals for each treatment group based on actual data rather than an extrapolated model (that is, excluding patients who were alive at the end of the trial). This analysis estimated a difference in mean overall survival of 1.92 months in favour of aflibercept.

3.8 The manufacturer found that aflibercept also prolonged progression-free survival compared with placebo; the difference in median progression-free survival was estimated to be 2.23 months when disease progression was assessed by an independent review committee (aflibercept 6.90 months, placebo 4.67 months, hazard ratio 0.758 [95% CI 0.661 to 0.869]). The manufacturer also provided an estimate of 1.74 months for median progression-free survival when local investigators determined disease progression. For response rate (complete and partial responses), the results favoured aflibercept, with an estimated response rate of 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.

3.9 The manufacturer performed pre-specified subgroup analyses according to the following:

- Baseline characteristics: presence of liver metastasis, location of primary tumour, number of metastatic organs (metastases in 1 organ only, or metastases in more than 1 organ), prior history of hypertension.
- Stratification variables: ECOG PS, prior bevacizumab treatment.
- Demographic characteristics: age (less than 65 years old, or 65 years or older), sex, race, geographical region.

The manufacturer focused on 2 subgroups in its submission: patients with liver metastases only (pre-specified), and a subgroup that excluded patients whose disease had relapsed 6 months or less after starting oxaliplatin-based adjuvant therapy (post hoc). The manufacturer stated that the subgroup of patients with liver metastases only was recognised as a relevant clinical subgroup for metastatic colorectal cancer in Cetuximab for the first-line treatment of metastatic colorectal cancer (NICE technology appraisal guidance 176). For the subgroup that excluded patients whose disease had relapsed 6 months or less after starting oxaliplatin-based adjuvant therapy, the manufacturer performed a post hoc analysis after the results of the VELOUR trial had been compiled. The manufacturer stated that 10% of patients in the trial had cancer that had relapsed within 6 months of starting oxaliplatin-based adjuvant therapy, which the manufacturer interpreted as reflecting patients with aggressive disease who would be unlikely to benefit from anti-vascular endothelial growth factor (VEGF) therapy.

3.10 For all the pre-specified subgroups, the manufacturer carried out an analysis of overall survival. It found no evidence of heterogeneity in

treatment effect (non-significant interaction test), except in the subgroup of patients with liver metastases only (p value for interaction was 0.0899, statistically significant at the 10% level). The hazard ratio for this subgroup was 0.649 (95.34% CI 0.492 to 0.855) compared with a hazard ratio of 0.868 (95.34% CI 0.742 to 1.015) in patients who had no liver metastases or in whom the cancer spread to the liver and other organs (estimates of survival times are academic in confidence). In response to a clarification request by the ERG, the manufacturer provided the difference in mean overall survival for the subgroup using actual, rather than extrapolated, data; this estimate is academic in confidence. In the post hoc subgroup analysis, which excluded patients whose disease had relapsed 6 months or less after starting oxaliplatin-based adjuvant therapy, the difference in median overall survival was estimated to be 1.9 months in favour of aflibercept. In this subgroup, the unadjusted hazard ratio was 0.78 (95% CI 0.68 to 0.90) compared with 1.09 (95% CI 0.70 to 1.69) in patients whose disease had relapsed 6 months or less after starting adjuvant therapy (p value for interaction 0.1265).

3.11 For progression-free survival, the manufacturer did not find a statistically significant subgroup effect except in patients with liver metastases only (interaction test was statistically significant at the 10% level). These results, and those of the subgroup that excluded patients whose disease had relapsed 6 months or less after starting oxaliplatin-based adjuvant therapy, are academic in confidence.

3.12 The incidence of adverse events of any grade (according to the Common Terminology Criteria for Adverse Events v3.0) was similar in the aflibercept and placebo groups of the VELOUR trial (99.2% and 97.9% respectively), but the incidence of some adverse events was considerably higher in the aflibercept group (for example,

41.4% of patients receiving aflibercept had hypertension [any grade] compared with 10.7% of those receiving placebo). Grade 3–4 adverse events were reported in 83.5% of patients in the aflibercept group and 62.5% of those in the placebo group. The grade 3–4 adverse events that occurred at least twice as frequently in the aflibercept group than in the placebo group, in order of decreasing relative incidence, were: hypertension (19.3% versus 1.5%), proteinuria (7.8% versus 1.2%), hand-foot syndrome (2.8% versus 0.5%), headache (1.6% versus 0.3%), arterial thromboembolic events (1.8% versus 0.5%), weight loss (2.6% versus 0.8%), stomatitis and ulceration (13.8% versus 5.0%), diarrhoea (19.3% versus 7.8%) and decreased platelet count (3.4% versus 1.6%). Typical anti-VEGF adverse reactions and adverse reactions associated with FOLFIRI were more common in the aflibercept group. The manufacturer indicated that most of the adverse events associated with aflibercept plus FOLFIRI were reversible and manageable using current clinical practice, although some (physical weakness, infections, diarrhoea and hypertension) led to permanent discontinuation of study treatment in 26.8% of patients receiving aflibercept compared with 12.1% of those receiving placebo. Furthermore, the European Public Assessment Report notes that more patients in the aflibercept than the placebo groups had their dose of FOLFIRI reduced or their treatment cycle delayed.

3.13 To further characterise the adverse events of aflibercept, the manufacturer performed a meta-analysis by pooling safety data from 3 RCTs (VELOUR, VITAL and VANILLA). The VITAL trial evaluated aflibercept plus docetaxel compared with placebo plus docetaxel in patients with non-small cell lung cancer and, in the VANILLA trial, patients with metastatic pancreatic cancer were

randomised to aflibercept plus gemcitabine or placebo plus gemcitabine. Overall, the meta-analysis included data from 2662 patients (1333 receiving aflibercept and 1329 receiving placebo). The analysis was framed so that risk ratios greater than 1 favoured placebo. The manufacturer found that, among patients treated with aflibercept, 0.4% and 0.5% had grade 4 hypertension and nephrotic syndrome respectively. It also found that adding aflibercept to concurrent chemotherapies did not increase the risk of venous thromboembolism, but it did increase the risk of grade 3–4 adverse reactions related to anti-VEGF therapy; the difference in this risk was statistically significant for hypertension (risk ratio [RR] 9.21, 95% CI 5.91 to 14.36), proteinuria (RR 8.37, 95% CI 4.37 to 16.06) and haemorrhage (RR 2.04, 95% CI 1.20 to 3.47). The incidence of adverse reactions typically associated with the background chemotherapy used in the 3 RCTs also increased with the addition of aflibercept, most notably for neutropenia (including neutropenic complications), various gastrointestinal toxicities and physical weakness.

3.14 Data on health-related quality of life were not collected in the VELOUR trial. The manufacturer conducted the 'mCRC utilities study', an observational, cross-sectional study to estimate utility values in patients with metastatic colorectal cancer who would be eligible for treatment with aflibercept plus FOLFIRI as per the licensed indication, or who had progressed to subsequent phases of the disease. The study took place in the Netherlands and the UK, and collected EQ-5D data. The manufacturer used these data as its main source to estimate health-related quality of life for the cost-effectiveness analysis.

ERG critique

3.15 The ERG stated that the manufacturer presented a well-conducted systematic review of clinical evidence, and used a search strategy that was unlikely to have missed any relevant studies. It also stated that the manufacturer included sufficient detail about the VELOUR trial and used appropriate criteria to assess the quality of the trial. The ERG noted, however, that the manufacturer provided minimal details of its meta-analysis of aflibercept's adverse events, and of the quality of the VITAL and VANILLA trials.

3.16 The ERG indicated that VELOUR was a good quality trial and directly related to the decision problem, and that the characteristics of patients at baseline and disease history were well balanced between the aflibercept and placebo groups. However, the ERG considered that patients in the trial were potentially fitter and younger than those seen in UK practice, and so patients in clinical practice may not achieve the level of benefit reported in the trial. The ERG highlighted the following dissimilarities between the VELOUR trial and clinical practice:

- In the UK, patients whose disease progresses after a break in treatment during intermittent first-line palliative chemotherapy are likely to be offered repeat treatment with the first-line chemotherapy regimen. If their disease progresses while receiving this treatment, or within 6 to 8 weeks of completing it, they would then move to second-line treatment. Although the manufacturer's submission does not state how many cycles of first-line oxaliplatin-based chemotherapy patients in the VELOUR trial received, the ERG indicated that the trial population may be healthier than patients in clinical practice who may have received several cycles of first-line treatment.

- Between 2007 and 2009, around 72% of patients diagnosed with colorectal cancer in the UK were aged 65 years or over. By contrast, in the VELOUR trial, only 33.5% of the aflibercept group and 38.9% of the placebo group were people aged 65 years or over.
- The proportion of patients with an ECOG PS of 2 in the VELOUR trial was 2.2%. According to the ERG's clinical adviser, this is lower than the proportion reported in other trials in the second-line setting, or in UK clinical practice.
- In the VELOUR trial, 42–44% of patients had metastasis in only 1 organ, which the ERG's clinical adviser considered higher than the proportion seen in clinical practice.

3.17 The ERG noted that the hazard ratios for overall survival by 6-month periods had wide confidence intervals at the later time points of the VELOUR trial because by this time many patients were no longer alive, leaving few patients at risk of dying (around 5% at 30 months). The ERG stated that wide confidence intervals reflect imprecise estimates, and that interpreting hazard ratios towards the end of the trial is highly uncertain, particularly at 30 months and 36 months.

3.18 To estimate mean overall survival using parametric analysis, the manufacturer assumed that the proportional hazards assumption does not hold (that is, it did not accept that the hazard ratio between the 2 treatment groups remained constant over time). The manufacturer stated that this was because the hazard ratios for overall survival decreased over time (treatment effect improved), and because the log-cumulative hazard plots were not parallel and crossed over one another. The ERG, conversely, considered that, while the hazard ratios decreased over time, they remained consistent with the proportional hazards assumption, although it

acknowledged that using a proportional hazards approach is subject to judgement. In addition, the ERG noted that the log-cumulative hazard plots were very close to parallel. The ERG stated that rejecting the proportional hazards assumption and assuming a continued separation of the overall survival curves is highly uncertain given that no data were available beyond 36 months follow-up, and particularly that the progression-free survival curves separate then converge at around 12 months. The ERG suggested that it would be reasonable to assume that the survival curves converge before 5 years (that is, there is no treatment effect after 5 years), in line with clinical experience in treating metastatic colorectal cancer.

3.19 The ERG noted that the estimate of mean overall survival varied considerably depending on the parametric function the manufacturer used, indicating that the manufacturer's estimates of the difference in mean overall survival (4.7 months) were not robust to the choice of distribution. The ERG requested from the manufacturer the mean estimates of overall survival for each treatment group, restricted to patients who had died before the end of the trial (that is, results based on actual data rather than an extrapolated model), which gave a difference of 1.92 months in favour of aflibercept. The ERG indicated that this figure is likely to be an underestimate given that it does not take into account the patients with long survival times.

3.20 The manufacturer used the log-logistic function to estimate mean overall survival, and it truncated the curves at 15 years. The ERG considered that 15 years is too long for the patient population under consideration because the treatment benefit is unlikely to extend beyond 5 years. The ERG requested that the manufacturer produce estimates with the analysis truncated at 5 years and

10 years. When the data were truncated at 5 years, the results from the different functions were more consistent with each other than when the data were truncated at 15 years. The ERG stated that it is unclear whether the mean based on extrapolating the curves and truncating the data at 5 years, or the restricted mean based on actual data, is more valid.

3.21 Progression-free survival in the VELOUR trial was a secondary end point assessed by an independent review committee. The ERG advised that independent review committees may miss symptoms other than tumour growth caused by disease progression, which may have an impact on treatment duration and associated costs. The ERG noted that, when the manufacturer explored in a sensitivity analysis disease progression determined by investigator assessment taking into account symptomatic deterioration (as would happen in clinical practice), aflibercept was found to extend median progression-free survival by 1.74 months.

3.22 The ERG stated that, while there was no evidence of a statistically significant interaction at the 5% level between treatment groups for most of the baseline patient characteristics, the results of the subgroup analyses suggested that patients with less advanced disease in the VELOUR trial (ECOG PS equal to 0, number of organs with metastasis less than or equal to 1, and patients with liver metastases only) may be more likely to benefit from treatment with aflibercept than those with more advanced cancer.

Cost-effectiveness evidence

3.23 The manufacturer did not identify any published economic evaluations relevant to the decision problem. It submitted a de novo economic model to establish the cost effectiveness of aflibercept in patients with metastatic colorectal cancer who are eligible for

second-line combination chemotherapy, and who were previously treated with an oxaliplatin-based regimen. The manufacturer performed subgroup analyses for patients with liver metastases only, and for a subgroup that excluded patients who had received oxaliplatin-based therapy in the adjuvant setting and whose disease relapsed within the following 6 months. The manufacturer conducted the analysis from the perspective of the NHS and personal social services and chose a time horizon of 15 years. It used a 2-week treatment cycle to reflect the treatment schedules of aflibercept and FOLFIRI, and applied a half-cycle correction. Costs and health effects were discounted at an annual rate of 3.5%.

3.24 The manufacturer developed a state-transition Markov cohort model simulating 3 states: stable disease, progressed disease and death. The manufacturer further split the stable-disease health state into sub-states of 'on second-line treatment' and 'discontinued second-line treatment' to distinguish between patients who receive second-line treatment until their disease progresses, and those who stop second-line treatment before their disease progresses. All simulated patients enter the model in the stable-disease health state and in the 'on second-line treatment' sub-state. Patients can then continue treatment and remain in the 'on second-line treatment' sub-state, or move to the 'discontinued second-line treatment' sub-state; they can instead move to the progressed-disease health state (and stop second-line treatment), or death. Patients cannot receive second-line treatment again once treatment is stopped, but they can receive further active therapy (systemic anticancer treatment, radiotherapy or surgery) or best supportive care. The manufacturer stated that the duration of second-line treatment in the model is based on the mean durations in the VELOUR trial to take into account dose delays or the

discontinuation of aflibercept or FOLFIRI (for patients who were in the aflibercept group), or FOLFIRI (for patients who were in the placebo group), as observed in the trial. The manufacturer modelled adverse events as events (rather than health states) and it applied a utility decrement (disutility) for each adverse event.

3.25 The manufacturer's model included parameters for overall survival, progression-free survival and time to discontinuing second-line treatment (before or after disease progression). To estimate the survival parameters, the manufacturer fitted alternative parametric functions (Weibull, log-normal, log-logistic and exponential) to observed Kaplan–Meier data from the VELOUR trial, and extrapolated the curves beyond the trial period for overall survival and time to discontinuing treatment, but not for progression-free survival, because the disease had progressed in all patients during the trial. In extrapolating those curves, the manufacturer assumed non-proportional hazards (that is, the hazard ratios between aflibercept plus FOLFIRI and FOLFIRI alone varied over time) so it modelled each treatment group separately. The manufacturer chose the base-case survival functions based on the results of statistical tests, visual inspection of the fit to the data and the clinical plausibility of the extrapolated portion of the curve. For overall survival, the manufacturer used the log-logistic function, and assumed that the survival benefit from treatment with aflibercept plus FOLFIRI increases relative to treatment with FOLFIRI alone until around 12 months after starting treatment, and then decreases over the 15-year time horizon, but does not cease at any point during the extrapolation period (that is, the overall survival curves start converging 12 months after starting treatment but never fully converge later in the extrapolation period). The manufacturer used the Weibull function for progression-free survival and time to

treatment discontinuation. The difference in mean progression-free survival estimated by the manufacturer was 1.2 months in favour of aflibercept [THIS ESTIMATE IS ACADEMIC IN CONFIDENCE]. Other parametric functions were explored in scenario analyses.

3.26 The manufacturer stated that the model predicted a median overall survival and a median progression-free survival similar to those from the VELOUR trial. The largest difference was for progression-free survival in the FOLFIRI group, which the model overestimated compared with the survival time observed in the trial.

3.27 Adverse events in the model included grade 3–4 adverse events that affected more than 5% of patients in the VELOUR trial, together with 6 rarer adverse events that the manufacturer's clinical advisory board considered important (gastrointestinal perforation, haemorrhage, febrile neutropenia, peripheral neuropathy, urinary tract infections and hand-foot syndrome). The subgroup analyses incorporated data specific to each subgroup.

3.28 The manufacturer applied utility values in the model from its 'mCRC utilities study', in which investigators assigned patients to 1 of the following 3 groups: patients with stable disease who are receiving second-line treatment, and patients who had previously received second-line treatment but stopped it because of an adverse event, or because their disease progressed. Because the sample size of the group of patients who had an adverse event and stopped treatment was very small, the manufacturer did not use the utility estimates from this group, and instead assumed that all patients with stable disease have the same utility, equal to the utility of patients with stable disease who are receiving second-line treatment. The manufacturer got descriptions of health states from patients using the EQ-5D system, and derived the utility weights by

applying UK valuation of health states estimated using the time trade-off method. The utility estimate used in the model for patients with progressed disease was 0.708 [THIS ESTIMATE IS ACADEMIC IN CONFIDENCE]. The manufacturer assumed that the utility in the progressed-disease health state is independent of time spent in the state. The manufacturer explained that, despite the age and health of patients, the utility values used in the model are relatively high because candidates for second-line chemotherapy must be fit enough to receive treatment.

3.29 The manufacturer also identified relevant utility studies from a systematic review of the literature. It did not use the values in those studies to source the model, but used them to compare the estimates from its utility study, and noted that they were reasonably consistent. The utility estimates in the literature that the manufacturer considered relevant ranged from 0.73 to 0.81 for stable disease, and from 0.68 to 0.69 for the progressed disease. One other study, Best et al. (2010), reported utility values of 0.51 for stable metastatic disease and 0.21 for progressed metastatic disease, but the manufacturer did not consider this study relevant because the population included patients receiving adjuvant chemotherapy and patients in remission.

3.30 The manufacturer got the disutilities associated with adverse events from the published literature, and supplemented these with clinical expert opinion. To calculate the average disutility per adverse event, the manufacturer assumed that an adverse event causes the same disutility regardless of the type of cancer. This gave an average disutility per adverse event of −0.0127 for patients receiving aflibercept plus FOLFIRI, and −0.0108 for those receiving FOLFIRI alone.

3.31 The costs of aflibercept plus FOLFIRI and FOLFIRI alone did not depend on the duration of second-line treatment in the model; the manufacturer calculated them separately based on data from the VELOUR trial to reflect the dose delays (for example, because of an adverse event) and dose reductions observed in the trial. It assumed that any unused drug in a vial was discarded (wasted) for aflibercept and irinotecan (a component of FOLFIRI), but explored in scenario analyses other possibilities to model drug wastage. The cost of aflibercept in the model took into account the patient access scheme discount.

3.32 To estimate costs of caring for people with metastatic colorectal cancer ('management costs' including supportive medications, clinician and nurse visits [hospital and community], imaging, laboratory tests, hospitalisations, palliative care, and personal and social care), the manufacturer conducted a retrospective observational study, and undertook a questionnaire-based survey of 6 UK clinical oncologists (both unpublished studies). In the observational study, the manufacturer collected resource-use data from patients who received oxaliplatin-based chemotherapy followed by FOLFIRI as second-line treatment, and used those data to estimate total management costs per 2-week cycle for different groups of patients (the manufacturer advised that every patient would eventually receive end-of-life care regardless of prior treatment, so it did not include resource use associated with end-of-life care in the model). The clinician survey aimed to gather data on community-based care, and on personal and social care. In this, the manufacturer elicited the average treatment practices of each oncologist to get data on managing patients with metastatic colorectal cancer. It also used the results of the survey, together with NHS reference costs, to estimate the costs associated with

adverse events. The manufacturer used mean resource use for adverse events, but median resource use for community-based care, and personal and social care. The cost of subsequent therapies that patients could receive after stopping second-line treatment or experiencing disease progression was calculated based on the manufacturer's study of resource use, and was assumed to be independent of the type of second-line treatment.

3.33 The manufacturer's deterministic base-case results estimated that the addition of aflibercept to FOLFIRI provides an additional 0.243 quality-adjusted life years (QALYs). This benefit is achieved with an additional cost of £8816, resulting in an incremental cost-effectiveness ratio (ICER) of £36,294 per QALY gained for aflibercept plus FOLFIRI compared with FOLFIRI alone.

3.34 The manufacturer presented deterministic sensitivity analyses in which it varied the 20 parameters with the largest impact on the ICER, one at a time. The results showed that the ICER is most sensitive to the parametric function chosen for overall survival, the utility value chosen for the progressed-disease health state, and the number of administrations assumed for second-line treatment drugs. The manufacturer explained that improving overall survival and progression-free survival increased incremental QALYs in favour of aflibercept, but also increased drug costs and the costs incurred from prolonged overall survival after disease progression.

3.35 The manufacturer carried out a probabilistic sensitivity analysis to summarise the uncertainty in the ICER. This showed that the probability of aflibercept plus FOLFIRI being cost effective when compared with FOLFIRI alone is less than 5% if the maximum acceptable ICER is £20,000 per QALY gained, and 22% at £30,000 per QALY gained.

3.36 The manufacturer investigated the structural uncertainty in the model by fitting alternative parametric functions for overall survival and progression-free survival, and by directly applying patient-level data from the VELOUR trial to model progression-free survival (given that disease had progressed in all patients during the trial). It also performed scenario analyses to test the sensitivity of the ICER to alternative assumptions around drug wastage. In these, it explored the possibility of no drug wastage, and of reducing the dose to the nearest number of whole vials for patients who would otherwise use less than 5% of the vial contents. The highest ICER from these analyses was £49,805 per QALY gained (using the Weibull function to model overall survival).

3.37 The manufacturer provided subgroup analyses to establish the cost effectiveness of aflibercept plus FOLFIRI compared with FOLFIRI alone in patients with liver metastases only, and in a subgroup that excluded those who had received oxaliplatin-based therapy in the adjuvant setting and whose disease had relapsed within the following 6 months. In comparison with the deterministic base-case ICER of £36,294 per QALY gained, the ICERs were £30,474 per QALY gained (incremental costs £10,974, incremental QALYs 0.360) and £32,480 per QALY gained (incremental costs £8573, incremental QALYs 0.264) respectively. At a maximum acceptable ICER of £30,000 per QALY gained, the probability of aflibercept plus FOLFIRI being cost effective compared with FOLFIRI alone in both subgroups is around 50% (numerical values not provided in the manufacturer's submission).

ERG critique

3.38 The ERG indicated that the manufacturer's economic evaluation is consistent with the NICE reference case. It noted that the modelled population is based on data from the VELOUR trial, which relate to

patients who appear fitter and younger than those seen in clinical practice. In exploratory sensitivity analyses, the ERG investigated the effect of treating a population that better reflects patients with metastatic colorectal cancer in the UK than the VELOUR trial by modelling an older population with a lower health-related quality of life.

3.39 The ERG considered that it is uncertain whether the hazard ratio for overall survival varies over time. The ERG reported that, when assuming in the manufacturer's model that the hazard ratio remains constant over time (that is, when applying the proportional hazards assumption), the ICER increased to £58,784 per QALY gained, with the difference being mainly driven by a reduction in incremental QALYs compared with the manufacturer's base case. The ERG considered that even this scenario may be relatively optimistic because the progression-free survival curves separate and then converge at around 12 months, suggesting that the hazard ratio could increase over time.

3.40 In its cost-effectiveness analysis, the manufacturer assumed that the survival benefit from treatment with aflibercept plus FOLFIRI initially increases relative to treatment with FOLFIRI alone until around 12 months after starting treatment, and then decreases over the rest of the time horizon, but does not cease at any point during the extrapolation period. The ERG noted that the difference in overall survival between aflibercept plus FOLFIRI and FOLFIRI alone decreases at a relatively slow rate after the initial 12 months and, importantly, suggests a continuing treatment effect on overall survival during the entire 15-year horizon. The ERG explained that extrapolating overall survival data from the VELOUR trial, in which the median follow-up time was just under 2 years, over a 15-year time horizon meant that the assumptions underpinning the

extrapolation are key to explaining the large differences between the observed median and the extrapolated mean estimates of overall survival. The ERG stressed that extrapolating the overall survival curves beyond the trial period is highly uncertain given that no data were available for more than 3 years' follow-up, and particularly that the progression-free survival curves separated and then converged at around 1 year. The ERG stated that the manufacturer did not explore this uncertainty sufficiently. Specifically, the manufacturer did not explore whether the risk of death in the aflibercept plus FOLFIRI and FOLFIRI alone groups could become the same from the point at which the trial ends (that is, the treatment effect of aflibercept plus FOLFIRI does not continue over the extrapolation period). In addition, it did not explore whether the overall survival curves for aflibercept plus FOLFIRI and FOLFIRI alone could converge over the extrapolation period (that is, the treatment effect of aflibercept plus FOLFIRI gradually decreases from the point at which the trial ends), similar to the convergence observed with progression-free survival (in this scenario the risk of death may be higher in the aflibercept plus FOLFIRI group during the extrapolation period than in the FOLFIRI alone group). The ERG explored these 2 scenarios in its exploratory analyses.

3.41 Regarding the utility estimates in the model, the ERG had concerns about the generalisability of the manufacturer's 'mCRC utilities study' because the study population appeared to be younger than UK patients, and the proportion of patients who had an ECOG PS of 2 was lower than that seen in UK clinical practice. Moreover, the ERG noted that the study was small, and produced counter-intuitive estimates in a subgroup analysis including UK patients only because the mean utility value for patients whose disease

progressed was higher than for those who had stable disease and received second-line treatment.

3.42 The ERG was concerned that the utility estimates used in the model from the manufacturer's utility study, as well as those reported in the literature, were high when compared with values used in previous appraisals of metastatic colorectal cancer, or with general UK population norms. The ERG was particularly concerned about the utility value in the model for patients whose disease had progressed. The ERG explained that, because the model predicts longer overall survival than progression-free survival, approximately three-quarters of absolute QALY increment is accrued after disease progression. Furthermore, the ERG stated that the manufacturer's assumption that utility in the progressed-disease health state is independent of time spent in the state is clinically implausible because patients' health-related quality of life decreases as disease progresses and patients get older.

3.43 The ERG identified an error in the manufacturer's model in how disutilities associated with adverse events were applied, which reduced the disutilities in the model. Correcting this error increased the manufacturer's base-case ICER from £36,294 to £37,834 per QALY gained. The ERG applied this correction in its exploratory analyses.

3.44 The costs of aflibercept plus FOLFIRI and FOLFIRI alone did not depend on the duration of second-line treatment in the model; the manufacturer calculated them separately based on data from the VELOUR trial to reflect the dose delays (for example, because of an adverse event) and dose reductions observed in the trial. The ERG stated that an alternative way to reflect dose delays and reductions would be to apply drug costs per administration

(including administration costs) directly to the proportion of patients in each health state, in line with how utility values are applied. Adjusting this increased the manufacturer's base-case ICER from £36,294 to £37,539 per QALY gained. The ERG applied this change in its exploratory analyses.

3.45 The manufacturer assumed that, because aflibercept is administered at the same time as FOLFIRI, no extra costs in terms of additional staff or inpatient admissions would be incurred. The ERG indicated that, even if given simultaneously, administering aflibercept involves preparing an additional infusion, which incurs an extra cost compared with FOLFIRI alone. The ERG highlighted that, in Cetuximab, bevacizumab and panitumumab for the treatment of metastatic colorectal cancer after first-line chemotherapy (NICE technology appraisal guidance 242), the pharmacy preparation of cetuximab and bevacizumab was estimated to be £15 per infusion. In addition, the ERG stated that, if aflibercept is given before or after FOLFIRI, instead of at the same time, administering aflibercept will include an additional hour of infusion time compared with administering FOLFIRI alone. The ERG noted that the model is sensitive to the assumptions underlying the administration costs of aflibercept plus FOLFIRI, and it explored these assumptions in sensitivity analyses.

3.46 Regarding resource use for community, and personal and social care, the manufacturer modelled the median estimate from its survey of clinical oncologists, instead of the mean. The ERG indicated that mean values are more commonly used in cost-effectiveness analyses, and that the use of medians may underestimate expected costs. The ERG noted that, when the manufacturer used the mean value in a sensitivity analysis, the base-case ICER increased from £36,294 to £41,222 per QALY

gained. The ERG stated that it is unclear in this case whether the median is a better estimate than the mean because there was a small number of survey responders (n=6) and the data were skewed. The ERG noted that the model is sensitive to this parameter and it further explored this in sensitivity analyses.

3.47 The ERG advised that the results of the analysis of the liver metastases only subgroup should be interpreted cautiously. Because the parametric curves for overall survival and progression-free survival were fitted independently for each treatment group based on data for this subgroup from the VELOUR trial, and the subgroup corresponded to approximately 25% of the trial population, the ERG highlighted that the analysis may not have been powered to demonstrate a difference in treatment effect in this subgroup. For the analysis of the subgroup that excluded adjuvant chemotherapy, the ERG indicated that this analysis was performed post hoc, and so its results may be biased. The ERG's clinical advisers also stated that patients who receive adjuvant chemotherapy and whose disease relapses quickly afterwards would not be treated differently to other patients in UK clinical practice.

ERG exploratory analyses

3.48 The ERG investigated the uncertainty around how the manufacturer had chosen to extrapolate overall survival by considering other scenarios for the magnitude and duration of the overall survival benefit associated with second-line treatments. The ERG modelled the following scenarios by assuming that:

- The risk of death in the aflibercept plus FOLFIRI and FOLFIRI alone groups becomes the same **30** months after starting treatment.

- The risk of death in aflibercept plus FOLFIRI and FOLFIRI alone groups becomes the same **36** months after starting treatment.

The ERG implemented the following scenarios to mimic the converging progression-free survival curves.

- The survival curves begin converging **30** months after starting treatment, and come together after a further **12** months, after which point the risk of death in both treatment groups becomes the same until the end of the time horizon.
- The survival curves begin converging **30** months after starting treatment, and come together after a further **18** months, after which point the risk of death in both treatment groups becomes the same until the end of the time horizon.
- The survival curves begin converging **36** months after starting treatment, and come together after a further **12** months, after which point the risk of death in both treatment groups becomes the same until the end of the time horizon.
- The survival curves begin converging **36** months after starting treatment, and come together after a further **18** months, after which point the risk of death in both treatment groups becomes the same until the end of the time horizon.

In all of the above scenarios, the ERG assumed that the treatment effect of aflibercept plus FOLFIRI continues until either 30 months or 36 months. The ERG chose these time points because it identified them as particularly uncertain from the hazard ratios for overall survival by 6-month periods presented by the manufacturer. When the ERG assumed that the risk of death in the aflibercept plus FOLFIRI and FOLFIRI alone groups becomes the same beyond the trial period, the ICERs were £45,570 and £42,718 per QALY gained for a treatment effect of aflibercept plus FOLFIRI

lasting until 30 months or 36 months respectively. In the scenario in which the ERG assumed that the survival curves begin converging 30 months or 36 months after starting treatment over a period of 12 months or 18 months, the ICERs ranged from £55,424 per QALY gained (when curves begin converging after 36 months over 18 months) to £66,377 per QALY gained (when curves begin converging after 30 months over 12 months). The ERG explained that, in this scenario, when the curves begin converging over 12 months, the magnitude of the additional survival benefit from treatment with aflibercept plus FOLFIRI is assumed to taper at a higher rate than when the curves begin converging over 18 months, and so convergence over 12 months results in higher ICERs.

3.49 To address its concerns about some of the parameters used in the manufacturer's base-case model, the ERG performed the following sensitivity analyses, varying 1 parameter at a time:

- Applying 2 alternative utility values for patients whose disease progressed: 0.21 from Best et al. (2010) and 0.60 from Bevacizumab and cetuximab for the treatment of metastatic colorectal cancer (NICE technology appraisal guidance 118). The ERG stated that the latter may better reflect the values reported in the literature.
- Including a cost for preparing an additional infusion of aflibercept, and a cost for an additional hour of infusion time for aflibercept plus FOLFIRI compared with administering FOLFIRI alone. For the preparation cost, the ERG applied a cost of £15, in line with NICE technology appraisal guidance 242 and, for the extra time for infusion, it applied £45, based on NHS reference costs. The ERG explored the impact of these 2 assumptions separately and jointly.

When the ERG used the lower utility values of 0.21 and 0.6, the ICER increased from £36,294 per QALY gained (base-case ICER) to £71,143 and £40,608 per QALY gained respectively. Including a cost for preparing an additional infusion of aflibercept, and a cost for an additional hour of infusion time for aflibercept plus FOLFIRI, together increased the ICER to £39,258 per QALY gained.

3.50 The ERG applied its preferred adjustments and model inputs to the manufacturer's base-case model (hereafter the 'ERG base case'). In this, the ERG corrected the error it identified in the manufacturer's model (section 3.43), and applied the acquisition and administration costs to all patients in the second-line treatment health state of the model (section 3.44). In addition, the ERG assumed that patients entered the model at the age of 70 years and accounted for the impact of age on health-related quality of life by applying a utility decrement for aging. The ICER resulting from the above 3 changes was £41,653 per QALY gained. The ERG then applied its preferred model inputs for the parameters it varied in one-way sensitivity analyses:

- an additional administration cost for aflibercept of £15
- mean instead of median resource use estimates (section 3.46).

The ERG applied the above with or without:

- a utility value of 0.60 for patients whose disease had progressed.

When the ERG applied the 0.60 utility value, the analysis gave an ICER of £54,368 per QALY gained for aflibercept plus FOLFIRI compared with FOLFIRI alone. Without this modification (that is, using the same value in the manufacturer's base case), the ICER was £47,965 per QALY gained.

3.51 The ERG presented deterministic results for the scenario analyses (section 3.48) within its base case, and using the utility value of 0.60 for patients whose disease had progressed. It presented results for the overall population, and separately for each subgroup the manufacturer had identified. When the ERG assumed that the risk of death in the aflibercept plus FOLFIRI and FOLFIRI alone groups becomes the same beyond the trial period, the ICERs were £66,506 and £62,894 per QALY gained for a treatment effect of aflibercept plus FOLFIRI lasting until 30 months or 36 months respectively. In the scenario in which the ERG assumed that the survival curves begin converging 30 months or 36 months after starting treatment over a period of 12 months or 18 months, the ICERs ranged from £78,226 per QALY gained (when the curves begin converging after 36 months over 18 months) to £92,089 per QALY gained (when the curves begin converging after 30 months over 12 months). The ERG found that, using median resource-use estimates from the manufacturer's survey of UK oncologists (that is, as per the manufacturer's base case), instead of mean, consistently decreased the ICERs for the scenario analyses within the ERG base case by approximately £5000 per QALY gained.

3.52 For the subgroup analyses combining the ERG's assumptions of overall survival and the ERG's alternative base case, the ICER for aflibercept plus FOLFIRI compared with FOLFIRI ranged from £46,576 to £58,257 per QALY gained for the liver metastases only subgroup, and from £57,224 to £80,187 per QALY gained for the subgroup that excluded patients who had received adjuvant oxaliplatin-based therapy and whose disease had relapsed within the following 6 months.

Manufacturer's response to consultation on the appraisal consultation document

3.53 To address the Committee's considerations of the evidence described in the appraisal consultation document, the manufacturer submitted a response to the consultation, which included:

- a revised patient access scheme discount (the details of which are commercial in confidence),
- utility data for the stable-disease state from an interim analysis of a phase III study (ASQoP), and
- proposed changes to parameters in the model considered by the Committee.

3.54 The ASQoP study was an international single-arm open-label phase III study. The primary objective of the study was to evaluate the safety of aflibercept in patients with metastatic colorectal cancer whose disease progressed following treatment with an oxaliplatin-based regimen. Its secondary objective was to establish health-related quality of life in this population. Because the study was not completed at the time of the second Committee meeting, the manufacturer provided interim results for mean EQ-5D utility values at baseline and after patients received 3 and 5 cycles of treatment. Data from this study were available for the stable-disease state only. The manufacturer derived a utility value of 0.78 for the stable-disease state by using a weighted average of the utility values for patients who received 3 and 5 cycles of treatment.

3.55 In its response, the manufacturer made the following comments on some of the parameters in the model originally considered by the Committee:

- The manufacturer considered that it was more clinically plausible to assume that the hazard ratio tapers to 1.0 after the end of the trial over a short period of time than to assume that the hazard ratio immediately changes to 1.0 at the end of the trial (the Committee's preferred extrapolation scenario).
- The manufacturer did not agree that the utility value chosen by the ERG for the progressed-disease state in its base case (0.6) was appropriate because it was based on a comparison with population 'norm' data that reflects the general population, which includes people with significant morbidities. The manufacturer stated that the utility value for progressed disease used in its original base case came from a relevant 'real-world' study that met the requirements of the NICE reference case. However, the manufacturer acknowledged that, according to clinical opinion, health-related quality of life declines sharply towards the end of life for patients with metastatic colorectal cancer.
- The manufacturer considered that assuming a starting age of 70 years in the model (as in the ERG base case) was too high according to available evidence and feedback from experts, and that a starting age of 60 years was more appropriate. The manufacturer provided the average age of patients with metastatic colorectal cancer receiving second-line treatment in 4 UK observational studies. It stated that these data were closer to the average age of patients in the VELOUR trial (60 years) than the average age used by the ERG (70 years).
- The manufacturer argued that the median value, rather than the mean value, from its survey of clinical oncologists was more appropriate for estimating resource use. This was because the data on the parameter for the number of visits received by a patient from a palliative care team contained a clear outlier, which had a significant impact on the ICERs. The manufacturer

further stated that the monthly cost of managing a patient whose disease had progressed used in NICE technology appraisal guidance 242 was closer to the median value than the mean.

3.56 The manufacturer revised its original base case by:

- applying a revised discount to the patient access scheme
- assuming that, 36 months after starting treatment, the hazard ratio for overall survival tapers to 1.0 over a 12 month period
- assuming that patients enter the model at the age of 60 years, and accounting for the impact of age on health-related quality of life by applying a utility decrement for aging
- updating the utility value of 0.78 for the stable-disease state from the ASQoP study
- correcting the disutilities associated with adverse events (section 3.43)
- including a cost of £15 for preparing an additional infusion of aflibercept, and a cost of £45 for additional administration time (£60 in total).

The manufacturer's deterministic results of the revised base case estimated that the addition of aflibercept to FOLFIRI would provide an additional 0.20 QALYs. This estimated benefit would cost an additional £8500, resulting in an estimated ICER of £42,242 per QALY gained for aflibercept plus FOLFIRI compared with FOLFIRI alone. The probabilistic ICER from this analysis was estimated to be £42,197 per QALY gained, and the probability of aflibercept plus FOLFIRI being cost effective when compared with FOLFIRI alone was around 10% if the maximum acceptable ICER was £30,000 per QALY gained, and 72% at £50,000 per QALY gained.

3.57 The manufacturer performed the following scenario analyses, in which it varied one parameter at a time:

- assuming that, **30** months after starting treatment, the hazard ratio for overall survival tapers to 1.0 over a 12 month period
- assuming that, **24** months after starting treatment, the hazard ratio for overall survival tapers to 1.0 over a 12 month period
- assuming that, **36** months after starting treatment, the hazard ratio changes to 1.0
- assuming that, **30** months after starting treatment, the hazard ratio changes to 1.0
- assuming that patients enter the model at the age of 65 years (while also applying a utility decrement for aging)
- applying the utility value for the stable-disease health state from the 'mCRC utilities study' (the value used in the manufacturer's original base case)
- applying a utility value of 0.3 during the last 2 months of life
- applying the mean value from its survey of clinical oncologists after excluding the outlier in the data on the number of visits received by a patient from a palliative care team
- applying the cost of managing disease progression used in NICE technology appraisal guidance 242.

The ICERs resulting from these scenario analyses ranged from £42,002 per QALY gained (when a utility value of 0.3 was applied during the last 2 months of life) to £47,246 per QALY gained (when the hazard ratio for overall survival begins tapering to 1.0 24 months after starting treatment over a 12-month period).

ERG critique of the manufacturer's revised base case

3.58 The ERG stated that the manufacturer's extrapolation of overall survival in its revised base case was not based on new data, and

so the ERG did not consider it any more plausible than the other scenarios previously presented to the Committee.

3.59 The ERG considered that the manufacturer's assumption of a 60-year age for starting treatment in the model was unrealistic, noting that 3 of the 4 observational studies provided by the manufacturer reported an average starting age of 63 years. However, the ERG also accepted that a starting age of 70 years may be high, and that an age of 65 years was a satisfactory compromise.

3.60 The ERG considered it appropriate for the manufacturer to have sourced the stable-disease utility value from the ASQoP study. However, the ERG argued that, because the manufacturer applied this value in the model for patients both on and off treatment, it would have been more appropriate to use the utility value of 0.77 for patients before they started treatment than the value for patients receiving treatment. The ERG indicated that the manufacturer's approach may have biased the utility value if patients receiving treatment were healthier than those who were not on treatment.

3.61 The ERG was concerned that, for the progressed-disease health state, the manufacturer continued to use the utility value from its 'mCRC utilities study', which the ERG considered high. Regarding the scenario analysis in which the manufacturer applied a utility value of 0.3 during the last 2 months of life, the ERG stated that this was not based on empirical evidence.

3.62 The ERG agreed that the estimate from the manufacturer's survey of UK oncologists included an outlier. It considered that using the mean value after excluding this outlier (as in the manufacturer's scenario analysis) was more appropriate than using the median.

3.63 To address remaining uncertainties, the ERG altered the manufacturer's revised base case by applying the utility value before treatment from the ASQoP study for the stable-disease state; the progressed-disease utility value of 0.6; and the mean resource use estimate from the manufacturer's survey of UK oncologists after excluding the potential outlier; and assuming patients start treatment at the age of 60 or 65 years. The ERG applied these changes together with each of the following extrapolation scenario:

- assuming a hazard ratio of 1.0 **30** months after starting treatment
- assuming a hazard ratio of 1.0 **36** months after starting treatment
- assuming that, **24** months after starting treatment, the hazard ratio tapers to 1.0 over 12 months
- assuming that, **30** months after starting treatment, the hazard ratio tapers to 1.0 over 12 months.

When the ERG assumed that patients start treatment at the age of 60 years, the resulting ICERs with the above scenarios were £54,243, £50,991, £55,139 and £51,296 per QALY gained respectively. When it assumed that patients start treatment at the age of 65 years, the ICERs were £54,890, £51,634, £55,791 and £51,941 per QALY gained respectively.

3.64 The ERG presented estimates of the difference in mean overall survival for different time horizons, while assuming a hazard ratio of 1.0 after 30 or 36 months. When the ERG set the time horizon to 5, 10 and 15 years, the differences in mean overall survival were 2.7–2.8, 3.2–3.5 and 3.4–3.7 months respectively.

3.65 Full details of all the evidence are in the manufacturer's submission and the ERG report.

4 Consideration of the evidence

4.1 The Appraisal Committee reviewed the data available on the clinical and cost effectiveness of aflibercept in combination with irinotecan and fluorouracil-based therapy, having considered evidence on the nature of metastatic colorectal cancer and the value placed on the benefits of aflibercept in combination with irinotecan and fluorouracil-based therapy by people with the condition, those who represent them, and clinical specialists. It also took into account the effective use of NHS resources.

4.2 The Committee heard from the clinical specialists and patient experts about the nature of the condition. It heard that metastatic colorectal cancer can be debilitating, can affect a person's ability to work and lead a normal life, and can lead to premature death. The Committee noted that the illness also brings about a burden on relatives and friends. The Committee understood that the course of the disease varies, with some people's health deteriorating quickly and others' slowly. The Committee heard from patient experts that quality of life in people with the disease may be bad when it is first diagnosed because patients are usually very weak and may have many metastases, but that with treatment the quality of life may improve, as may the ability to work and socialise. The Committee understood that, in clinical practice, disease progression (in patients who already have metastatic disease) would be detected using radiological imaging, although symptoms would also be taken into account. It heard from patient experts that disease progression usually affects quality of life, but it may take a long time before it affects daily activities. The Committee heard further from patient

experts that, although people would appreciate small extensions of life, they value quality of life more than length of life. The Committee noted that treatment is generally associated with unpleasant side effects, particularly high blood pressure and diarrhoea and that, while some people may be willing to tolerate the side effects, others may not. The Committee understood that clinicians are now more experienced in managing these side effects and 'optimising' treatment, although it heard from patient experts that the treatments for the side effects may themselves have side effects.

4.3 The Committee discussed the management of metastatic colorectal cancer. It heard from clinical specialists that the current treatment options for this patient population are limited, and that treatment is determined individually. The Committee was aware that, in Colorectal cancer: the diagnosis and management of colorectal cancer (NICE clinical guideline 131), NICE recommends, as second-line treatment options, single-agent irinotecan or folinic acid/5-fluorouracil/irinotecan (FOLFIRI) after first-line folinic acid/5-fluorouracil/oxaliplatin (FOLFOX), and single-agent irinotecan after first-line capecitabine/oxaliplatin (XELOX). The Committee heard from the clinical specialists that resecting tumours surgically may be a treatment option in some patients with metastatic disease, noting that systemic therapy can make resection possible in some patients. The Committee understood that the proportion of patients with metastatic colorectal cancer who survive over 5 years has increased because of successful tumour resection. The Committee noted that patients consider biologic therapies such as aflibercept to improve quality of life compared with chemotherapy.

Clinical effectiveness

4.4 The Committee considered the evidence on the clinical effectiveness of aflibercept, noting that it was derived from the VELOUR trial. The Committee agreed that the VELOUR trial was of good quality and directly relevant to the decision problem; however, it considered that the trial had limitations. The Committee, echoing comments from a patient expert, would have liked the manufacturer to have collected and presented trial data relating to health-related quality of life. The Committee would also have liked the manufacturer to have followed and presented event data for all patients after the end of the trial as defined. The Committee discussed the initial Evidence Review Group's (ERG) concern that patients in the VELOUR trial appeared to have been fitter and younger than those seen in UK clinical practice. The Committee also heard from clinical specialists that the disease and demographic characteristics seen in patients in the VELOUR trial differed from those for patients treated in UK clinical practice; however, evidence from the VELOUR trial showed that response to treatment does not vary across patient groups. The Committee was aware that the studies provided by the manufacturer in response to consultation suggested that patients receiving second-line treatment for metastatic colorectal cancer in the UK were somewhat older than those in the VELOUR trial. The Committee agreed that the patient population in the VELOUR trial was otherwise reasonably representative of patients seen in the UK and therefore concluded that the results of the VELOUR trial are generalisable to UK clinical practice.

4.5 The Committee discussed the results for overall survival, the primary end point of the VELOUR trial. The Committee noted that, in its submission and in response to clarification requests by the

ERG, the manufacturer produced a range of estimates for the difference in overall survival between the aflibercept and placebo groups of the trial. The Committee considered that the data observed directly from the trial were sufficiently mature at the study cut-off date to establish median overall survival, and agreed that the difference in median overall survival of 1.44 months reflects a statistically significant but clinically small benefit. The Committee noted that the restricted mean difference of 1.92 months (based on the unlikely and conservative assumption that all remaining patients die immediately at the end of the trial) was higher than the median. The Committee therefore concluded that the difference in median overall survival is likely to underestimate the mean survival benefit of aflibercept.

4.6 The Committee discussed the estimated mean survival benefit of 4.7 months derived from extrapolation, in light of the trial data. The Committee noted that, to estimate this benefit, the manufacturer extrapolated the survival curves from a trial with a median follow-up of just under 2 years up to 15 years. The Committee noted the comments received in the ACD consultation, and agreed to explain its concerns over the survival extrapolation in more detail. The Committee noted the marked difference between the estimated mean survival benefit of 4.7 months and that of 1.44 months based on median values of overall survival. The Committee understood that the manufacturer considered the extrapolated mean value to represent the magnitude of the clinical benefit of aflibercept better than the median because there were a few patients who experienced a sustained survival benefit from treatment with aflibercept. Although the Committee agreed that a small proportion of patients, with as yet undefined characteristics, appeared to derive greater benefit from aflibercept than most patients in the

trial, it agreed that extrapolating the survival curves over 15 years could result in highly uncertain estimates for overall survival. The Committee therefore discussed whether the manufacturer's estimates of mean overall survival were robust. It noted that, during the trial's maximum 3-year follow-up period, 66% of patients in the aflibercept group and 75% of those in the placebo group had died, and a proportion had been censored (data academic in confidence), reducing the number of patients at risk of dying to 47 patients (3.8%) at 30 months and 1 patient (0.1%) at 36 months. The Committee heard from the ERG that, because of this, the hazard ratios for overall survival for patients with follow-up nearing 36 months had wide confidence intervals, reflecting imprecise and uncertain estimates. The Committee noted that, at 36 months follow-up, the Kaplan–Meier curves indicated that approximately 17% of patients randomised to aflibercept were alive but that this 17% represented only 1 patient remaining uncensored and at risk of dying, which considerably increased the uncertainty around the long-term effect on survival. The Committee was aware that the cut-off date for the trial was 07 February 2011, and that the manufacturer had continued to follow-up patients for overall survival. The Committee met 18 months after this date, but the manufacturer did not present the Committee with follow-up data to support its extrapolation. The Committee appreciated that estimating mean overall survival often requires extrapolating beyond a trial period, but considered that the manufacturer's extrapolation of overall survival from a population with very few patients at risk of dying after 30 months follow-up, over a further 12 years, was associated with great uncertainty. The Committee concluded that, given the data from the VELOUR trial, the true mean overall survival benefit is likely to be closer to the median

estimate of 1.44 months than the manufacturer's mean estimate of 4.7 months.

4.7 The Committee discussed the mean survival benefit of 4.7 months in light of the different parametric functions used by the manufacturer to estimate overall survival. The Committee was aware that, to estimate this benefit, the manufacturer used the log-logistic function, which it considered to provide the best fit to the observed data, and extrapolated the survival curves over 15 years. The Committee noted that the estimates using other parametric functions ranged from 3.0 months (with the Weibull function) to 5.3 months (with the log-normal function). The Committee discussed which extrapolation period could be considered appropriate to estimate mean overall survival, in view of the life expectancy of patients with metastatic colorectal cancer in clinical practice. The Committee was aware that extrapolation periods should reflect the time in which all patients will have died, but that a longer than 5-year survival for patients with metastatic colorectal cancer is very unusual. It was also aware that, with surgical resection of liver metastases, survival can increase, but that a very small proportion of patients in the VELOUR trial had surgical resection of liver metastases (data designated as academic in confidence), and the Committee was not presented with information about their survival. The Committee also considered survival statistics from the US cancer registry Surveillance, Epidemiology, and End Results (SEER), which showed that 6.9% of patients with metastatic colorectal cancer survive for 5 years. However, because this registry included patients who had received multiple lines of therapy, including surgical resections of tumours and therapies that may not have been considered established NHS practice, the Committee did not consider the data from the SEER registry to be a

reasonable proxy for the life expectancy of the population specified in the marketing authorisation of aflibercept. The Committee agreed that a shorter extrapolation period better reflected the natural history of the disease at this stage, and yet accounted for patients who derived greater benefit from aflibercept than most patients in the VELOUR trial. It considered the robustness of the mean overall survival benefit, obtained using the log-logistic function, of 3 months (5 years extrapolation time), 4.7 months (15 years extrapolation time) and 6.6 months (without truncating the survival curves). The Committee was concerned that the log-logistic function had a very 'heavy tail' (that is, a high probability of getting large values at the end of the time horizon) compared with other parametric functions, and that this is likely to have led to an overestimate of the survival benefit of aflibercept. The Committee was also concerned that the manufacturer did not characterise the uncertainty around any of the estimates. In summary, the Committee concluded that, because of the uncertainties around the survival extrapolation, the actual trial data and the life expectancy of patients at this stage of the disease, extrapolating overall survival with the log-logistic function over 15 years did not provide a plausible mean overall survival benefit.

4.8 The Committee considered the relationship between progression-free survival and overall survival from the VELOUR trial. The Committee was aware that the manufacturer used disease progression assessed by an independent review committee in its base case. The difference in median progression-free survival between aflibercept and placebo using this methods was 2.23 months, which was a higher value than when disease progression was determined by investigator assessment (1.74 months) and higher than the mean

progression-free survival (see section 3.25; numerical value designated as confidential by the manufacturer).The Committee considered the shapes of the Kaplan–Meier curves (reflecting the trial data) for overall survival and for progression-free survival. It noted that the curves continued to diverge during the trial period for overall survival, whereas, for progression-free survival, the curves initially diverged but then converged at around 12 months, reflecting almost the same rate of progression for patients randomised to aflibercept or placebo from that time onward. The Committee heard from the clinical specialists that, because the overall survival curves continued to separate for both patients who had or had not stopped treatment, the survival curves might reflect a disease-modifying effect in that aflibercept might have altered the natural course of the disease whereby, despite the disease progressing, patients lived longer even after treatment stopped (that is, survival post disease progression was increased more than progression-free survival). The Committee discussed how the disease-modifying effect could be explained clinically. It heard that aflibercept may have delayed death by shrinking tumours, and so extended the period before the tumour grew again. However, the Committee was not presented with evidence that tumours had shrunk, and was aware that the disease had progressed in all patients during the trial. The Committee agreed that there was no robust evidence to make firm conclusions about the likely cause of the different shapes of the overall survival and progression-free survival curves, and that the magnitude of progression-free survival depended on the method used to calculate it.

4.9 The Committee considered that the subgroup analysis presented by the manufacturer for patients with liver metastases only compared with metastases not confined to the liver. The Committee

noted that, in this subgroup, there was a statistically significant interaction test at the 10% level. The Committee was aware that the 10% significance level was less specific (that is, a higher chance of a positive finding) than the more conventional 5% level. The Committee agreed that there is no evidence to suggest that aflibercept would be more effective in patients with liver metastases only than in patients with metastases confined to other organs. The Committee was aware that patients with liver metastases only are more likely to be considered for surgical resection of the metastases and possibly live longer than those with widespread metastases. The Committee therefore discussed whether aflibercept can make liver metastases operable in patients with metastatic colorectal cancer. It noted that only a very small minority of patients in the VELOUR trial proceeded to have surgical resection of liver metastases after treatment with aflibercept. The Committee heard from the clinical specialist that, in approximately 20–30% of patients who have surgery to remove liver metastases, metastatic colorectal cancer can be cured. The Committee, however, was not presented with evidence about rates of resection and cure with aflibercept in the subgroup of patients with liver metastases only. The Committee also considered that resecting liver metastases to achieve a cure was more appropriate in the first-line setting than in the second-line setting. Furthermore, it heard from the manufacturer that the modelling of the subgroup did not include the costs of surgical resection. The Committee concluded that it would be appropriate to include this cost and that including it is likely to affect the incremental cost-effectiveness ratio (ICER). The Committee agreed that, given the lack of evidence, aflibercept cannot be considered an effective treatment option to make liver metastases resectable. The Committee therefore concluded that this subgroup should not be considered further.

4.10 The Committee considered the subgroup that excluded patients who had received oxaliplatin-based therapy in the adjuvant setting and whose disease had relapsed within the following 6 months. The Committee heard from the clinical specialists that, in clinical practice, patients in this subgroup would not be treated differently to the overall trial population. In addition, the Committee noted that the manufacturer acknowledged that the analysis for this subgroup was planned after the trial results had been compiled (post hoc), and that the test for interaction did not show that the treatment effect in this subgroup differed from the effect in the rest of the trial population. The Committee therefore concluded that it did not need to consider further the subgroup that excluded patients whose disease had relapsed 6 months or less after starting oxaliplatin-based adjuvant therapy.

4.11 The Committee discussed the adverse events associated with aflibercept. The Committee noted that more patients in the aflibercept group (27%) stopped treatment because of adverse events than in the placebo group (12%). The Committee also noted that adding aflibercept to FOLFIRI increased the adverse events typically associated with FOLFIRI, most notably neutropenia, although it heard from clinical specialists that neutropenia would not routinely be treated in clinical practice. The Committee heard from the manufacturer that the dose of FOLFIRI used in the trial was higher than the dose that is routinely used in clinical practice and might have caused some of the adverse events. The Committee was also aware that aflibercept increased the risk of hypertension, as would other anti-vascular endothelial growth factor therapies. The Committee concluded that treatment with aflibercept plus FOLFIRI was associated with a considerable burden of

adverse effects, but that, being a new treatment, less is known about its adverse effects profile than other available treatments.

Cost effectiveness

4.12 The Committee considered the structure of the model submitted by the manufacturer, and how it captured the main aspects of the condition. The Committee noted that the manufacturer chose to split the stable-disease health state into 2 sub-states to capture the costs and health benefits for patients with stable disease who either receive second-line treatment with aflibercept plus FOLFIRI or FOLFIRI alone, or who have stable disease but have stopped second-line treatment for reasons other than disease progression. The Committee heard from the ERG that the manufacturer applied the same utility value to the 2 sub-states of the stable-disease health state. It further heard that the acquisition and administration costs of second-line treatments in the model did not depend on the proportion of patients in each state, and that they were calculated outside the model. The Committee noted that the costs and quality-adjusted life years (QALYs) in the stable-disease health state were not specific to the 2 sub-states ('on second-line treatment' and 'post second-line treatment'). The Committee concluded that overall the model adhered to the NICE reference case for assessing cost effectiveness.

4.13 The Committee discussed whether or not the 15-year time horizon used by the manufacturer in the model was appropriate. The Committee appreciated that that the choice of the time horizon is a sensitive parameter in the model given the uncertainty associated with extrapolating overall survival. The Committee was aware that the time horizon should be sufficiently long to capture all the costs and health benefits in the full population (that is, a lifetime horizon

should be used). The Committee therefore concluded that a time horizon of 15 years was, in principle, appropriate because all patients are likely to have died by 15 years; however, the Committee agreed that, when the time horizon is much longer than the trial duration, and the life expectancy of most patients, it is particularly important to explore the assumptions underlying how overall survival is extrapolated.

4.14 The Committee discussed the manufacturer's assumptions for extrapolating overall survival in the model, and the alternative assumptions considered by the ERG in its exploratory analyses. The Committee noted that the manufacturer assumed that the survival benefit from treatment with aflibercept plus FOLFIRI increases relative to treatment with FOLFIRI alone until around 12 months after starting treatment, and then decreases over the 15-year time horizon, but that the hazard ratio never reaches 1.0 (that is, a patient previously randomised to aflibercept will always have a lower risk of dying, even if not receiving aflibercept, relative to a patient previously randomised to placebo). The Committee noted that the ERG explored 2 alternative scenarios:

- the first assumed that the risk of death becomes the same in both treatment groups at the point at which the trial ends and continues to be the same for the remainder of the time horizon period (that is, the hazard ratio becomes 1.0 after 3 years)
- the second assumed that the overall survival curves for aflibercept plus FOLFIRI and FOLFIRI alone converge over the time horizon (that is, the hazard ratio gradually increases from the end of the trial until the survival curves come together, then the hazard ratio becomes 1.0 thereafter).

The Committee understood that, in the ERG's second scenario, patients receiving aflibercept plus FOLFIRI need to have a higher risk of death than patients receiving FOLFIRI alone (that is, the hazard ratio may be greater than 1.0) for the curves to converge. The Committee considered that the manufacturer's assumption that the treatment benefit continues beyond the trial period and until 15 years is highly uncertain given that most patients had died during the 3-year follow-up period of the trial. The Committee considered that the ERG's analysis that allows the hazard ratio to become greater than 1.0 could be considered implausible. The Committee agreed that the ERG's first scenario, which assumes equal risk of death for all patients beyond the trial period (hazard ratio equals 1.0), represents an acceptable compromise between the 2 extremes of assuming continuing treatment effect (manufacturer's base case) and allowing for a reversed treatment effect (ERG's second scenario). The Committee noted that, in response to consultation, the manufacturer implemented a new scenario in its revised base case in which the hazard ratio begins to taper to 1.0 36 months after starting treatment, over a 12-month period. The Committee agreed that all scenarios to extrapolate overall survival were associated with some degree of uncertainty. As the manufacturer had not provided the follow-up data from the trial on which to validate its new approach, the Committee chose the ERG's first scenario.

4.15 The Committee considered the estimates of health-related quality of life used in the manufacturer's model, noting that it would have preferred these data to have been collected from the VELOUR trial. The Committee was aware that the manufacturer got the utility value for the stable-disease state from the 'mCRC utilities study' and revised it after consultation to a value derived from the ASQoP

study because the data from this study were new, and not because the Committee questioned the validity of the original value. The Committee noted that the ERG preferred another value from the ASQoP study for the stable-disease state but, because the difference between the manufacturer's revised value (0.78) and the ERG's preferred value (0.77) was small and likely to have a negligible impact on the ICER, the Committee concluded that either value could be considered appropriate.

4.16 The Committee discussed the appropriate utility value for the progressed-disease state in the model, noting that, because approximately three-quarters of the QALY gain in the model was accrued after disease progression, the model is highly sensitive to this parameter. The Committee, having noted the mean and median time to disease progression in the manufacturer's utility study, considered that the utility value chosen by the manufacturer for the progressed-disease state did not reflect the entire duration of progressed disease but only early progressed disease, and so was likely to be an overestimate (see section 3.28; data academic in confidence). The Committee was aware of the participation bias associated with studies of this nature. Furthermore, it heard from the ERG that the manufacturer's study was small, and produced counter-intuitive estimates in 1 subgroup analysis. The Committee was aware that, although the manufacturer stated that the data queries noted in its submission had been resolved, the manufacturer had yet to submit the study for peer-reviewed publication. The Committee was aware that, in its base case, the ERG used an alternative lower value of 0.60, which had been used in NICE technology appraisal guidance 118, and that the ERG considered this value to represent a reasonable balance of the utility values for progressed disease used in other NICE guidance,

which ranged from 0.21 to 0.69. The Committee was aware that the utility value of 0.69 used in NICE technology appraisal guidance 242 for progressed disease was based on patients who had lived long enough to receive more courses of chemotherapy than patients in the VELOUR trial, and so likely reflected patients with a better health state. The Committee agreed that no utility values for progressed disease were universally accepted as valid, but that it would be important that the utility value reflected the entire progressed-disease state. The Committee was aware that the quality of life for patients with metastatic colorectal cancer deteriorates relatively slowly other than during the last few months, when it may deteriorate faster, and that exploring a utility value of 0.3 during the last 2 months of life was a reasonable attempt by the manufacturer to address this. The Committee also agreed that adjusting the utility values for age was appropriate to reflect the natural deterioration in health-related quality of life in patients with the disease. The Committee concluded that the most plausible utility value for the progressed-disease health state would lie between the manufacturer's and the ERG's estimate.

4.17 The Committee discussed the costs of administering aflibercept plus FOLFIRI in the model, noting that the manufacturer assumed no extra cost for administering aflibercept in its original model. The Committee was aware that aflibercept would normally be prepared in a sterile compartment, and would therefore incur an extra cost; the Committee estimated that this cost is likely to be higher than the £15 used by the ERG. The Committee was also aware that the marketing authorisation for aflibercept stipulates that aflibercept should be administered over 1 hour before the infusion with FOLFIRI, but that the cost for an additional hour of infusion time (£45) was not included in the ERG base case. The Committee

acknowledged that the manufacturer's revised base case accounted for the extra preparation cost and the cost for an additional infusion time for aflibercept.

4.18 The Committee noted that, in response to consultation, the manufacturer had provided data showing that the average age of patients treated in the NHS with second-line chemotherapy for metastatic colorectal cancer was 60 years in 1 study and 63 years in 3 others. The Committee agreed that the 70-year age of starting treatment, as initially assumed by the ERG in its base case, was therefore too high. It concluded that an age between 60 and 65 years is more appropriate.

4.19 The Committee discussed the costs in the model derived from the manufacturer's survey of clinical oncologists about community-based care, and personal and social care. The Committee noted that this study was small and therefore associated with uncertainty, and did not provide evidence that the oncologists in the survey were representative of practitioners in the UK. The Committee noted that the manufacturer's model incorporated median estimates from the survey because the responses from clinicians on 1 parameter (the number of visits received by a patient from a palliative care team) included an outlier, whereas the ERG argued that the mean was more appropriate. The Committee agreed that, if the sample of clinicians was appropriately homogenous and reflected similar practices, the distribution of the data collected from the survey would be largely uniform, and it would be more appropriate to use the mean rather than the median. The Committee noted that, although the manufacturer continued to use the median value in its revised base case, it presented a scenario analysis that incorporated the mean after excluding the outlier, an approach that the ERG considered

appropriate. Although the Committee agreed that mean values should normally be used to estimate resource use and costs, it concluded that, in this instance, using the mean after excluding the outlier could be considered appropriate.

4.20 The Committee discussed whether aflibercept should be considered an innovative treatment. The Committee acknowledged that aflibercept represented a novel recombinant fusion protein. However, the Committee concluded that all benefits of a substantial nature relating to treatment with aflibercept plus FOLFIRI had been captured in the QALY calculation.

4.21 The Committee considered supplementary advice from NICE that should be taken into account when appraising treatments that may extend the life of patients with a short life expectancy and that are licensed for indications that affect small numbers of people with incurable illnesses. For this advice to be applied, all the following criteria must be met.

- The treatment is indicated for patients with a short life expectancy, normally less than 24 months.
- There is sufficient evidence to indicate that the treatment offers an extension to life, normally of at least an additional 3 months, compared with current NHS treatment.
- The treatment is licensed or otherwise indicated for small patient populations.

In addition, when taking these criteria into account, the Committee must be persuaded that the estimates of the extension to life are robust and that the assumptions used in the reference case of the economic modelling are plausible, objective and robust.

4.22 The Committee considered the criterion for short life expectancy and the evidence for life expectancy in this group of patients. The Committee noted the overall survival estimates presented by the manufacturer from the VELOUR trial with the observed median survival in the placebo group of VELOUR of 12.1 months and the estimated mean overall survival of 18.1 months. The Committee also noted the ERG's preferred estimate of 10.5 months from the literature. The Committee concluded that patients receiving current standard NHS treatment would have an expected survival of less than 24 months from the point at which they would be considered for second-line therapy and that therefore the criterion for short life expectancy was fulfilled in this appraisal.

4.23 The Committee considered the criterion that the treatment is licensed or otherwise indicated for small patient populations. The Committee noted the manufacturer's suggestion that approximately 4000 patients in England and Wales would receive second-line treatment for metastatic colorectal cancer. The Committee was concerned that aflibercept holds a marketing authorisation for treatment of a much larger population with neovascular (wet) age-related macular degeneration, but that this was a different formulation of aflibercept marketed by another company. The Committee understood that when one technology is marketed by different companies (for different indications, using different brands), these should not be added for the purpose of establishing the cumulative population to be considered in the context of life-extending treatments at the end of life, and that therefore the criterion for a small population size was fulfilled in this appraisal.

4.24 The Committee considered the criterion that treatment offers an extension to life of normally at least an additional 3 months. The Committee noted the comments received in consultation on the

ACD, and agreed to explain its concerns over the magnitude of the mean survival benefit more fully. The Committee noted that, based on the number of patients who had died during the trial (70.4%), 50% of those who received aflibercept lived for up to 1.44 months longer than people who received placebo, and acknowledged the difficulty in finding robust mean overall survival data considering the issues with the extrapolation carried out (see section 4.7). The Committee noted that, in response to consultation, the manufacturer pointed out that the original base-case model, using the Committee's preferred assumption to extrapolate overall survival (section 4.14), predicted that aflibercept would extend life by 3.4-3.7 months. The Committee discussed whether the estimates for mean overall survival produced by the model were robust indicators of what overall survival benefit can be seen in clinical practice, noting that all of the extrapolation assumptions were associated with great uncertainty. The Committee was aware that the longer the time horizon, the greater the influence of the 'tails' of the extrapolation curves, which define the difference in mean overall survival between the treatment arms, and to which the model is highly sensitive. The Committee agreed that, although there is a rationale for a 15-year time horizon in order to capture the very small number of patients who might have very prolonged survival, this introduced considerable uncertainty, and produced implausible results given that the extrapolation was based on a population with a small number of patients still at risk of dying beyond 30 months. Although the use of a 15-year time horizon is, in principle, appropriate, when extrapolating the relative benefit is associated with uncertainty, the Committee considered it appropriate to consider shorter time horizons as a means to explore the uncertainty. The Committee noted that, when the model time horizon was shortened to 5 years, the difference in mean

overall survival decreased to 2.7–2.8 months. The Committee was mindful that, when there is quantitative evidence that a treatment offers a 3-month life extension, it must also be persuaded that the estimates of life extension are robust and that the assumptions used in the reference case of the economic modelling are plausible, objective and robust. The Committee agreed that, given the considerable uncertainty around extrapolating overall survival and its implementation in the modelling, it is important to take into account what has actually been observed in the trial (see section 4.6 and 4.8) and in the absence of other evidence (including the lack of data on survival after 07 February 2011 from the VELOUR trial in support of the overall survival claims), the Committee was not satisfied that the estimates from fitting parametric functions to Kaplan–Meier data or those produced by the model were sufficiently robust to accept that the 3-month life extension criterion is fulfilled. The Committee therefore concluded that aflibercept did not meet the criteria for an end-of-life therapy as defined by NICE.

4.25 The Committee noted that, in its response to consultation on the appraisal consultation document, the manufacturer pointed out that, in NICE technology appraisal guidance 242, the Committee had considered a modelled survival benefit of 2.7–3.2 months to show sufficient evidence for a 3-month survival benefit for panitumumab. The Committee was aware that, in judging whether panitumumab met the criterion for life extension, the Committee for NICE technology appraisal 242 had taken into consideration the difficulty in accommodating the cross-over in the panitumumab trials and that the mean progression-free survival benefit for panitumumab was similar to that for cetuximab, and that the latter resulted in an overall survival benefit of 4.7 months. Therefore, it had considered

that there was sufficient evidence to indicate that panitumumab offers an extension to life of approximately 3 months.

4.26 The Committee discussed the ICERs for aflibercept in combination with irinotecan and fluorouracil-based therapy for metastatic colorectal cancer based on the revised analyses provided in response to consultation. The Committee agreed that the cost-effectiveness analysis should assume equal risk of death for all patients beyond the trial period, and that the starting age of the modelled cohort should be between 60 and 65 years. The Committee noted that the manufacturer's ICER closest to these assumptions was £44,000 per QALY gained (for age 60), but would increase for the higher age bracket, if the mean value was used from the manufacturer' survey of clinical oncologists after removing the outlier and if an extrapolation function with a less heavy tail had been used. Because the manufacturer's ICERs incorporated a utility value for progressed disease deemed by the Committee to be high, the Committee considered the ICER produced by the ERG using the Committee's preferred assumptions, but which used a utility value for progressed disease of 0.6. The Committee noted that this was approximately £51,000 per QALY gained and would be higher if an extrapolation function with a less heavy tail had been used. The Committee therefore concluded that the most plausible ICER was higher than the normally acceptable maximum ICER range of £20,000–30,000 per QALY gained, and that aflibercept in combination with irinotecan and fluorouracil-based therapy could not be considered a cost-effective use of NHS resources for patients with metastatic colorectal cancer.

4.27 The Committee noted the comments received during consultation on the appraisal consultation document that some patients appeared to gain particular benefit ('a bimodal distribution'), and

which stressed the importance of offering only certain patients aflibercept. The Committee was aware that there is currently no established method in clinical practice to identify patients with metastatic colorectal cancer who could particularly benefit from treatment, and it was not presented with evidence on how these patients could be selected for treatment with aflibercept. The Committee was aware that, as a post-authorisation commitment to the European Medicines Agency, the manufacturer initiated a biomarker program encompassing 3 studies to help select patients who may be more likely to benefit. The Committee agreed that the results of these studies would be useful for a future review of this appraisal.

Summary of Appraisal Committee's key conclusions

TAXXX	**Appraisal title:**	**Section**
Key conclusion		
Aflibercept in combination with irinotecan and fluorouracil-based therapy is not recommended within its marketing authorisation for treating metastatic colorectal cancer that is resistant to or has progressed after an oxaliplatin-containing regimen.		1.1
Given the considerable uncertainty around extrapolating overall survival and its implementation within the model, and in the absence of other evidence, the Committee was not satisfied that the estimates from fitting parametric functions to Kaplan–Meier data or those produced by the model were sufficiently robust to accept that the 3-month life extension criterion is fulfilled.		4.24
The Committee concluded that the most plausible ICER was higher than the normally acceptable maximum ICER range of £20,000–30,000 per QALY gained.		4.26

Current practice		
Clinical need of patients, including the availability of alternative treatments	The Committee noted that the current treatment options for patients with metastatic colorectal cancer are limited, and that treatment is determined individually. The Committee heard that resecting tumours surgically may be a treatment option in some patients with metastatic disease, noting that systemic therapy can make resection possible in some patients.	4.3
The technology		
Proposed benefits of the technology How innovative is the technology in its potential to make a significant and substantial impact on health-related benefits?	The Committee noted that patients consider biologic therapies such as aflibercept to improve quality of life compared with chemotherapy. The Committee heard that, because the overall survival curves continued to separate for both patients who had or had not stopped treatment, the survival curves might reflect a disease modifying effect in that aflibercept might have altered the natural course of the disease whereby, despite the disease progressing, patients lived longer even after treatment stopped. The Committee agreed that there was no robust evidence to make firm conclusions about the likely cause of the different shapes of the overall survival and progression-free survival curves. The Committee acknowledged that aflibercept represented a novel recombinant fusion protein. However, the Committee concluded that all benefits of a substantial nature relating to treatment with aflibercept plus FOLFIRI had been captured in the QALY calculation	4.3 4.8 4.20
What is the position of the treatment in the pathway of care for the condition?	Aflibercept in combination with FOLFIRI has a UK marketing authorisation 'for the treatment of adults with metastatic colorectal cancer that is resistant to or has progressed after an oxaliplatin containing regimen'.	2.1
Adverse reactions	The Committee concluded that treatment with aflibercept plus FOLFIRI was associated with a considerable burden of adverse effects, but that, being a new treatment, less is known about its adverse effects profile than other available treatments.	4.11

Evidence for clinical effectiveness		
Availability, nature and quality of evidence	The Committee noted that the evidence on the clinical effectiveness of aflibercept was derived from the VELOUR trial. The Committee agreed that the VELOUR trial was of good quality and directly relevant to the decision problem. However, the Committee would have liked the manufacturer to have collected and presented trial data relating to health-related quality of life, and would have liked the manufacturer to have followed and presented event data for all patients after the end of the trial as defined. The Committee concluded that the results from the VELOUR trial are generalisable to UK clinical practice.	4.4
Relevance to general clinical practice in the NHS	No specific Committee considerations on the relevance to general clinical practice in the NHS.	

Uncertainties generated by the evidence	The Committee noted that, to estimate aflibercept's mean survival benefit of 4.7 months, the manufacturer extrapolated the survival curves from a trial with a median follow-up of just under 2 years up to 15 years. Although the Committee agreed that a small proportion of patients, with as yet undefined characteristics, appeared to derive greater benefit from aflibercept than most patients in the trial, it considered that the manufacturer's extrapolation of overall survival based on a population with a very few patients at risk of dying after 30 months follow-up over a further 12 years was associated with great uncertainty.	4.6
	The Committee was aware that the manufacturer estimated the mean survival benefit of 4.7 months by fitting the log-logistic function to the observed data, and extrapolating the survival curves over 15 years. The Committee noted that the estimates using other parametric functions ranged from 3.0–5.3 months. The Committee was aware that a longer than 5-year survival for patients with metastatic colorectal cancer is very unusual. Having considered the estimates obtained using different parametric functions and extrapolation periods, the Committee was concerned that the log-logistic function had a very 'heavy tail', and that this is likely to have overestimated the survival benefit of aflibercept. The Committee was also concerned that the manufacturer did not characterise the uncertainty around any of the estimates. The Committee concluded that extrapolating overall survival with the log-logistic function over 15 years did not provide a plausible mean overall survival benefit.	4.7

Are there any clinically relevant subgroups for which there is evidence of differential effectiveness?	The Committee agreed that there was no evidence to suggest that aflibercept would be more effective in patients with liver metastases only than in patients with metastases confined to other organs. The Committee was not presented with evidence about rates of resection and cure with aflibercept in the subgroup of patients with liver metastases only. The Committee therefore agreed that aflibercept cannot be considered an effective treatment option to make liver metastases resectable, concluding that this subgroup should not be considered further. The Committee heard from the clinical specialists that, in clinical practice, patients who had received oxaliplatin-based therapy in the adjuvant setting and relapsed within the following 6 months would not be treated differently to the overall trial population. In addition, the Committee noted that the analysis for this subgroup was planned after the trial results had been compiled (post hoc), and that the test for interaction did not show that the treatment effect in this subgroup differed from the effect in the rest of the trial population. The Committee therefore concluded that it did not need to consider further the subgroup that excluded patients whose disease had relapsed 6 months or less after starting oxaliplatin-based adjuvant therapy.	4.9 4.10
Estimate of the size of the clinical effectiveness including strength of supporting evidence	The Committee agreed that the difference in median overall survival of 1.44 months reflects a statistically significant but clinically small benefit. The Committee concluded that, given the data from the VELOUR trial, the true mean overall survival benefit is likely to be closer to the median estimate of 1.44 months than the manufacturer's mean estimate of 4.7 months.	4.5 4.6
Evidence for cost effectiveness		
Availability and nature of evidence	The Committee concluded that overall the manufacturer's model adhered to the NICE reference case for assessing cost effectiveness.	4.12

Uncertainties around and plausibility of assumptions and inputs in the economic model	The Committee considered that the manufacturer's assumption that the treatment benefit continues beyond the trial period and until 15 years is highly uncertain given that most patients had died during the 3-year follow-up period of the trial. The Committee considered that the ERG's analysis that allows the hazard ratio to become greater than 1.0 could be considered implausible. The Committee agreed that the ERG's scenario, which assumes equal risk of death for all patients beyond the trial period (hazard ratio equals 1.0), represents an acceptable compromise between the 2 extremes of assuming continuing treatment effect (manufacturer's base case) and allowing for a reversed treatment effect (ERG's second scenario). The Committee noted that, in response to consultation, the manufacturer implemented a new scenario in its revised base case. The Committee agreed that all scenarios to extrapolate overall survival were associated with some degree of uncertainty. As the manufacturer had not provided the follow-up data from the trial on which to validate its new approach, the Committee chose the ERG's first scenario, which assumes equal risk of death for all patients beyond the trial period (hazard ratio equals 1.0).	4.14

Incorporation of health-related quality-of-life benefits and utility values Have any potential significant and substantial health-related benefits been identified that were not included in the economic model, and how have they been considered?	The Committee was aware that the manufacturer got the utility value for the stable-disease state from the 'mCRC utilities study' and revised it after consultation to a value derived from the ASQoP. The Committee noted that the ERG preferred another value from the ASQoP study for the stable-disease state. The Committee concluded that either value could be considered appropriate. The Committee considered that the utility value chosen by the manufacturer for the progressed-disease state did not reflect the entire duration of progressed disease but only early progressed disease, and so was likely to be an overestimate. The Committee was aware that, in its base case, the ERG used an alternative lower value of 0.60, which had been used in NICE technology appraisal guidance 118. The Committee agreed that no utility values for progressed disease were universally accepted as valid, but that it would be important that the utility value reflected the entire progressed-disease state. The Committee also agreed that adjusting the utility values for age was appropriate. The Committee concluded that the most plausible utility value for the progressed-disease health state would lie between the manufacturer's and the ERG's estimate. The Committee concluded that all benefits of a substantial nature relating to treatment with aflibercept plus FOLFIRI had been captured in the QALY calculation.	4.15 4.16 4.20
Are there specific groups of people for whom the technology is particularly cost effective?	Having considered the clinical evidence presented by the manufacturer for the 2 subgroups, the Committee concluded that it did not need to consider the cost effectiveness of the technology for any of the subgroups.	4.9 4.10

What are the key drivers of cost effectiveness?	The Committee considered the robustness of the mean overall survival benefit, obtained using the log-logistic function, of 3 months (5 years extrapolation time), 4.7 months (15 years extrapolation time) and 6.6 months (without truncating the survival curves. The Committee was aware that the longer the time horizon, the greater the influence of the 'tails' of the extrapolation curves, which define the difference in mean overall survival between the treatment arms, and to which the model is highly sensitive. The Committee noted that, because approximately three-quarters of the QALY gain in the model was accrued after disease progression, the model is highly sensitive to utility value for the progressed-disease state in the model.	4.7 4.24 4.16
Most likely cost-effectiveness estimate (given as an ICER)	The Committee noted that the manufacturer's ICER closest to its preferred assumptions was £44,000 per QALY gained (for age 60), but would increase for the higher age bracket, if the mean value was used from the manufacturer' survey of clinical oncologists after removing the outlier and if an extrapolation function with a less heavy tail had been used. Because the manufacturer's ICERs incorporated a utility value for progressed disease deemed by the Committee to be high, the Committee considered the ICER produced by the ERG using the Committee's preferred assumptions, but which used a utility value for progressed disease of 0.6. The Committee noted that this was approximately £51,000 per QALY gained and would be higher if an extrapolation function with a less heavy tail had been used. The Committee therefore concluded that the most plausible ICER was higher than the normally acceptable maximum ICER range of £20,000–30,000 per QALY gained.	4.26
Additional factors taken into account		
Patient access schemes (PPRS)	The manufacturer of aflibercept (Sanofi) has agreed a patient access scheme with the Department of Health that makes aflibercept available with a discount. The size of the discount is commercial in confidence.	2.3

End-of-life considerations	The Committee agreed that, given the considerable uncertainty around extrapolating overall survival and its implementation within the model, it is important to take into account what has actually been observed in the trial (see section 4.6 and 4.8) and, in the absence of other evidence (including the lack of data on survival after 07 February 2011 from the VELOUR trial in support of the overall survival claims), the Committee was not satisfied that the estimates from fitting parametric functions to Kaplan–Meier data or those produced by the model were sufficiently robust to accept that the 3-month life extension criterion is fulfilled. The Committee therefore concluded that aflibercept did not meet the criteria for an end-of-life therapy as defined by NICE.	4.24
Equalities considerations and social value judgements	No equality issues relevant to the Committee's recommendations were raised.	

5 Implementation

5.1 The Department of Health and the manufacturer have agreed that aflibercept will be available to the NHS with a patient access scheme which makes aflibercept available with a discount. The size of the discount is commercial in confidence. It is the responsibility of the manufacturer to communicate details of the discount to the relevant NHS organisations. Any enquiries from NHS organisations about the patient access scheme should be directed to [NICE to add details at time of publication]

5.2 NICE has developed tools [link to www.nice.org.uk/guidance/TAXXX] to help organisations put this guidance into practice (listed below). [NICE to amend list as needed at time of publication]

- Slides highlighting key messages for local discussion.

- Costing template and report to estimate the national and local savings and costs associated with implementation.
- Implementation advice on how to put the guidance into practice and national initiatives that support this locally.
- A costing statement explaining the resource impact of this guidance.
- Audit support for monitoring local practice.

6 Related NICE guidance

Details are correct at the time of publication. Further information is available on the NICE website.

Published

- Colorectal cancer: the diagnosis and management of colorectal cancer. NICE clinical guideline 131 (2011).
- Cetuximab (monotherapy or combination chemotherapy), bevacizumab (in combination with non-oxaliplatin chemotherapy) and panitumumab (monotherapy) for the treatment of metastatic colorectal cancer after first-line chemotherapy (review of TA150 and part review of TA118). NICE technology appraisal guidance 242 (2012).
- Panitumumab in combination with chemotherapy for the treatment of metastatic colorectal cancer (terminated appraisal). NICE technology appraisal 240 (2011).

7 Review of guidance

7.1 The guidance on this technology will be considered for review in August 2016. The Guidance Executive will decide whether the technology should be reviewed based on information gathered by NICE, and in consultation with consultees and commentators.

Amanda Adler

Chair, Appraisal Committee

June 2013

8 Appraisal Committee members, guideline representatives and NICE project team

8.1 *Appraisal Committee members*

The Appraisal Committees are standing advisory committees of NICE. Members are appointed for a 3-year term. A list of the Committee members who took part in the discussions for this appraisal appears below. There are 4 Appraisal Committees, each with a chair and vice chair. Each Appraisal Committee meets once a month, except in December when there are no meetings. Each Committee considers its own list of technologies, and ongoing topics are not moved between Committees.

Committee members are asked to declare any interests in the technology to be appraised. If it is considered there is a conflict of interest, the member is excluded from participating further in that appraisal.

The minutes of each Appraisal Committee meeting, which include the names of the members who attended and their declarations of interests, are posted on the NICE website.

Dr Amanda Adler (Chair)
Consultant Physician, Addenbrooke's Hospital

Professor Ken Stein (Vice Chair)
Professor of Public Health, University of Exeter Medical School

Dr Ray Armstrong
Consultant Rheumatologist, Southampton General Hospital

Dr Jeff Aronson
Reader in Clinical Pharmacology, University Department of Primary Health Care, University of Oxford

Professor John Cairns

Professor of Health Economics Public Health and Policy, London School of Hygiene and Tropical Medicine

Mark Chapman

Health Economics and Market Access Manager, Medtronic UK

Professor Fergus Gleeson

Consultant Radiologist, Churchill Hospital, Oxford

Robert Hinchliffe

HEFCE Clinical Senior Lecturer in Vascular Surgery and Honorary Consultant Vascular Surgeon, St George's Vascular Institute

Professor Daniel Hochhauser

Consultant in Medical Oncology, UCL Cancer Institute

Dr Neil Iosson

General Practitioner

Anne Joshua

Associate Director of Pharmacy, NHS Direct

Dr Rebecca Kearney

Clinical Lecturer, University of Warwick

Professor Ruairidh Milne

Director of Strategy and Development and Director for Public Health Research at the National Institute for Health Research (NIHR) Evaluation, Trials and Studies Coordinating Centre at the University of Southampton

Dr Elizabeth Murray

Reader in Primary Care, University College London

Dr Peter Norrie

Principal Lecturer in Nursing, DeMontfort University

Dr Sanjeev Patel

Consultant Physician & Senior Lecturer in Rheumatology, St Helier University Hospital

Dr John Pounsford

Consultant Physician, Frenchay Hospital, Bristol

Alun Roebuck

Consultant Nurse in Critical and Acute Care, United Lincolnshire NHS Trust

Roderick Smith

Chief Finance Officer, Coastal West Sussex Clinical Commissioning Group

Cliff Snelling

Lay Member

Professor Andrew Stevens

Professor of Public Health, Department of Public Health and Epidemiology, University of Birmingham

Dr Nicky Welton

Senior Lecturer in Biostatistics/Health Technology Assessment, University of Bristol

8.2 *NICE project team*

Each technology appraisal is assigned to a team consisting of 1 or more health technology analysts (who act as technical leads for the appraisal), a technical adviser and a project manager.

Ahmed Elsada

Technical Lead(s)

Joanna Richardson

Technical Adviser

Jeremy Powell

Project Manager

9 Sources of evidence considered by the Committee

A. The Evidence Review Group (ERG) report for this appraisal was prepared by the CRD and CHE Technology Assessment Group, University of York:

- Wade R, Duarte A et al. Aflibercept in combination with irinotecan and fluorouracil-based therapy for the treatment of metastatic colorectal cancer which has progressed following prior oxaliplatin-based chemotherapy, April 2013

B. The following organisations accepted the invitation to participate in this appraisal as consultees and commentators. They were invited to comment on the draft scope, the ERG report and the appraisal consultation document (ACD). Organisations listed in I were also invited to make written submissions. Organisations listed in II and III had the opportunity to give their expert views. Organisations listed in I, II and III also have the opportunity to appeal against the final appraisal determination.

I. Manufacturer/sponsor:

- Sanofi

II. Professional/specialist and patient/carer groups:

- Beating Bowel Cancer
- Cancer Research UK
- Royal College of Nursing

III. Other consultees:

- Department of Health
- Welsh Government

IV. Commentator organisations (did not provide written evidence and without the right of appeal):

- Commissioning Support Appraisals Service
- Department of Health, Social Services and Public Safety for Northern Ireland
- Healthcare Improvement Scotland
- Medicines and Healthcare products Regulatory Agency
- Pfizer
- Roche

C. The following individuals were selected from clinical specialist and patient expert nominations from the consultees and commentators. They gave their expert personal view on aflibercept by attending the initial Committee discussion and providing written evidence to the Committee. They were also invited to comment on the ACD.

- Dr Richard Adams, Senior Lecturer and Consultant Oncologist, Velindre Cancer Centre and Cardiff University, nominated by Sanofi - clinical specialist
- Jacqueline Fraser, nominated by Beating Bowel Cancer - patient expert
- Helen Minnery, nominated by Beating Bowel Cancer - patient expert

D. Representatives from the following manufacturer/sponsor attended Committee meetings. They contributed only when asked by the Committee chair to clarify specific issues and comment on factual accuracy.

- Sanofi