The CA184-024 trial was a large good quality RCT with reliable results, however this trial does not provide direct evidence for the effectiveness of ipilimumab 3 mg/kg monotherapy (without maintenance therapy) compared with DTIC, vemurafenib or dabrafenib for the treatment of previously untreated, advanced melanoma patients.

The MDX010-08 trial was under powered to detect a statistically significant difference in overall survival,[6] and the BREAK-3[8] and BRIM-3[7] trials of dabrafenib and vemurafenib were reasonably good quality, although they were both open-label and had a relatively short duration of follow-up.

The quality of the non-RCT studies was generally adequate;[9-11] however, there is a high risk of bias associated with these study designs, relative to an RCT. The manufacturer compared the results of the non-RCT studies with the DTIC arm of the CA184-024 trial,[5] however this is problematic due to the differences in patient characteristics and study design. In addition, some of the RCT data were included in the pooled dataset, in addition to being presented separately in the MS. The ERG does not consider that these studies provide additional relevant reliable data.

The ERG does not consider the indirect comparisons or mixed treatment comparison of ipilimumab plus DTIC, vemurafenib and dabrafenib to be appropriate because of differences in patients' baseline characteristics between the trials (see below).

The ERG has three main concerns with the argument presented in the MS.

Comparability of trial populations

The main ipilimumab trial population differs from the population included in the vemurafenib and dabrafenib trials (which are restricted to patients who are BRAF V600 mutation positive); these trials were used in the indirect and mixed treatment comparisons. There were differences in patients' baseline characteristics between the BRIM-3 and BREAK-3 trials and the CA184-024 trial; patients in the BRAF mutation-positive trials had a worse prognosis (differences in the proportion of patients with elevated serum lactate dehydrogenase levels and M1c disease).

Assumption of equivalent clinical effectiveness of ipilimumab monotherapy versus ipilimumab plus DTIC

The MS cites study MDX010-08 as evidence for the equivalence of ipilimumab monotherapy and ipilimumab plus DTIC.[6] However, the results of this very small (n=64 patients) trial suggest that ipilimumab plus DTIC may be more effective than ipilimumab alone. Therefore, there is no conclusive evidence that ipilimumab monotherapy is as clinically effective as ipilimumab plus DTIC.

BSC in the mutation-positive population. The modelled pathway for DTIC in both sub-populations was first-line DTIC followed by second-line ipilimumab and third-line BSC. The modelled pathway for vemurafenib was first-line vemurafenib followed by second-line ipilimumab and third-line BSC in the mutation-positive population.

The assessment of cost-effectiveness was based on comparative effectiveness evidence on the efficacy of ipilimumab 10mg/kg in addition to DTIC therapy compared with DTIC therapy alone from the CA184-024 trial,[5] and the separate, independent arm from the BRIM-3 trial was used for vemurafenib.[7] The transitions between lines of treatment and the proportion of patients who die in each weekly cycle of the model were informed by extrapolated empirical data on overall survival (OS) and progression-free survival (PFS) from the clinical trials. For second-line treatments, OS was based on first-line survival curves but adjusted downwards to account for poorer outcomes on second-line therapy using a constant proportional hazard derived from expected life years of second-line ipilimumab (TA268).[1] The duration of response to second-line treatments was based on the number of pre-progression life years for second-line ipilimumab. Third-line treatment was assumed to be BSC, which consisted of a proportion of patients on 'no treatment' and a proportion on commonly prescribed chemotherapy drugs, including DTIC. The OS for patients who survived to receive third-line BSC was assumed to be the same as those on first-line treatment independent of progression status. Quality of life was quantified by applying utility weights based on time-to-death, derived from the EORTC QLQ-C30 questionnaire data from CA184-024[5] in order to estimate quality-adjusted life years (QALYs). Costs were assessed from an NHS and personal and social services perspective and incorporated acquisition, administration and monitoring costs of the alternative regimens, adverse events and other supportive care and terminal care costs associated with the management of progressed disease.

The manufacturer presented separate pair-wise comparisons of the incremental cost-effectiveness ratio (ICER) for ipilimumab with DTIC in the mutation-negative (£16,958 per QALY) and –positive populations (£31,559 per QALY) and ipilimumab with vemurafenib in the mutation-positive population (ipilimumab dominates vemurafenib with less costs and more QALYs). The manufacturer also presented results based on a pooled chemotherapy-naïve dataset (from CA184-004,[13] CA184-022,[12] MDX010-08,[6] MDX010-020[18] clinical trials), which showed similar and consistent ICERs to the base-case analysis. The manufacturer presented a series of one-way deterministic sensitivity analyses and scenario analyses to assess the impact of uncertainty around key input variables and assumptions on the ICER estimates. The results of these indicated that the base-case ICER estimates were most sensitive to the assumptions used to model overall survival of ipilimumab and DTIC, time spent on second-line treatments, and time spent on first-line treatment with ipilimumab relative to

DTIC. The manufacturer's probabilistic sensitivity analysis concluded that the probability of ipilimumab being the most cost-effective treatment compared with DTIC was 96% at a willingness to pay threshold of £50,000 per additional QALY and 40% at a threshold of £30,000 per QALY, while it was close to 100% for the comparison with vemurafenib.

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

The ERG considers the manufacturer's base-case ICERs to be both highly uncertain (largely due to the assumptions employed by the manufacturer for overall survival on second- and third-line treatments) and overly optimistic towards ipilimumab. The main driver of cost-effectiveness for the comparison of ipilimumab with DTIC and vemurafenib is the assumptions around the sequencing of treatments. The ERG considers the approach used by the manufacturer to estimate OS on second-line treatments and third-line BSC as arbitrary. The manufacturer switches between different OS curves at different lines of treatment such that overall survival on BSC is often better than on second-line treatment. The method used by the manufacturer is not consistent or clinically plausible, and substantially favours the cost-effectiveness of ipilimumab. The cost-effectiveness of ipilimumab is achieved by assuming that patients who progress to BSC following first-line ipilimumab achieve the long-term benefits of sustained overall survival from ipilimumab first-line. However, for the comparator treatment (DTIC or vemurafenib) patients who receive ipilimumab second-line cannot continue to receive the long-term benefits of sustained overall survival from ipilimumab second-line when they progress to BSC, instead they return to the lower overall survival of the first-line treatment (DTIC or vemurafenib).

The model also makes a number of important assumptions about the relative efficacy of the treatments. The manufacturer's submission assumes that the absence of comparative efficacy evidence for the licensed dose of ipilimumab 3mg/kg in previously untreated advanced melanoma implies equal efficacy to ipilimumab 10mg/kg, and that concomitant DTIC therapy with ipilimumab 10mg/kg in the CA184-024 trial[5] implies equal efficacy to ipilimumab 10mg/kg monotherapy. Furthermore, the model assumes that the use of separate, independent arms from different clinical trials (i.e. breaking randomisation) is sufficient to establish the relative efficacy of ipilimumab with vemurafenib. The model also employs higher acquisition costs of ipilimumab per weekly cycle when treated second-line compared with first-line, which ultimately disadvantages the comparator treatment where ipilimumab is received second-line. The ERG considers that these critical assumptions have not been robustly justified by the manufacturer and are not clearly supported by the existing evidence.

The ERG believes that the only plausible model structure on which to base the estimates of cost-effectiveness of ipilimumab is to consider first-line treatments only, which essentially uses a conventional three-state model based on pre-progression, post-progression, and death, with BSC

DTIC in the mutation-negative population increased from £16,958 per QALY to between £18,833 and £56,486 per QALY. For the same comparison in the mutation-positive population, the manufacturer's base-case ICER increased from £31,559 to between £71,335 and £941,091 per QALY. For the comparison with vemurafenib in the mutation-positive population, the ICER of ipilimumab which dominated vemurafenib in the manufacturer's base-case analysis moved to the south-west quadrant of the cost-effectiveness plane resulting in an ICER for vemurafenib of between £27,180 and £84,980. The ERG concluded that the assumptions used by the manufacturer to model the sequencing of treatments were arbitrary and that the only plausible model structure on which to base the estimates of cost-effectiveness was to consider first-line treatments only. The ERG, however, noted that even with the use of a simpler model structure, the estimates of cost-effectiveness remained highly sensitive to the assumptions used to model overall survival on BSC. Assuming that ipilimumab and its comparator treatment moved to the same OS curve upon progression to BSC resulted in an increase in the three-state model ICER of £42,449 to £123,676 per QALY for the comparison of ipilimumab with DTIC. Under this same assumption (i.e. same OS curves for BSC), the three-state model ICER for the comparison of ipilimumab with vemurafenib (£28,980 per QALY) became favourable to vemurafenib such that it dominated ipilimumab.

The assumptions surrounding the manufacturer's comparative effectiveness estimates of ipilimumab with DTIC and vemurafenib also had a major impact on the cost-effectiveness results. In order to estimate overall survival for the licensed dose of ipilimumab 3mg/kg, the ERG used a pooled data analysis comparing OS profiles of ipilimumab 3mg/kg (MDX010-20; CA184-022) and 10mg/kg (CA184-007; CA184-008; CA184-022) for a mixed (both previously treated and untreated) patient population, which was presented to the ERG following Points of Clarification to the manufacturer.[14] The ERG estimated an implied hazard ratio for ipilimumab 3mg/kg based on this comparison, which was then used to adjust the overall survival curve for ipilimumab 10mg/kg. Taking the three-state model as the ERG's preferred base-case, the impact of the adjustment on the estimate of cost-effectiveness was significant. For the comparison of ipilimumab with DTIC in the mutation-positive or -negative population, the adjustment for 3mg/kg yielded considerably lower QALYs (1.56 QALYs compared with the base-case of 2.35 QALYs) and resulted in a multi-fold ICER of £331,091 per QALY. A similar analysis was undertaken to assess the impact of concomitant DTIC in the ipilimumab 10mg/kg + DTIC arm of the CA184-024 trial. The effect on the ICER was also remarkable (increased to £674,144 per QALY for the same comparison).

The ERG noted a discrepancy between the acquisition costs of ipilimumab per weekly cycle for first- (▇▇▇) and second-line (▇▇▇) treatment. The ERG explored the implications of this cost difference on the estimates of cost-effectiveness by allocating equal acquisition costs per cycle

4 Clinical Effectiveness

This section contains a critique of the methods of the review of clinical effectiveness data, followed by a description and critique of the trials included in the review, including a summary of their quality and results. The ERG's conclusions on the clinical effectiveness of ipilimumab for the treatment of previously untreated advanced malignant melanoma are presented at the end of this section.

4.1 Critique of the methods of review(s)

The MS described a systematic review evaluating the clinical efficacy and tolerability of ipilimumab and comparator therapies for the treatment of adult patients with previously untreated advanced malignant melanoma. No direct evidence was identified which evaluated the effectiveness of ipilimumab monotherapy, at its licensed dose of 3 mg/kg, compared with DTIC, vemurafenib or dabrafenib for the treatment of previously untreated, advanced melanoma patients. The evidence presented in the MS comprised of four RCTs identified in the systematic review, along with additional non-RCT evidence selected by the manufacturer.

4.1.1 Search strategy

The MS described the systematic review search strategies used to identify relevant clinical effectiveness studies on the use of ipilimumab and comparator therapies for the treatment of adult patients with previously untreated advanced malignant melanoma. The search strategies were described in the main body of the submission, and full details were provided in the Appendices.

The electronic databases MEDLINE, MEDLINE In Process, EMBASE, the Cochrane Library (including the Cochrane Database of Systematic Reviews, the Cochrane Central Register of Controlled Trials, the Health Technology Assessment Database and the Database of Abstracts of Reviews of Effects (DARE)), CINAHL and ClinicalTrials.gov were searched in May 2013. In addition, the 2009 to 2012 conference proceedings of the American Society of Clinical Oncology (ASCO), the European Society for Medical Oncology (ESMO) and the Society for Melanoma Research (SMR) were searched on 9th May 2013. The manufacturer attempted to search the conference proceedings of the European Association of Dermato-Oncologists (EADO) and Perspectives in Melanoma (PiM), however the abstracts could not be accessed as abstracts are only made available to meeting delegates, so could not be searched. The reference lists of previous systematic reviews and clinical guidelines identified were hand searched to identify any additional relevant studies, and unpublished data from Bristol-Myers Squibb were reviewed for relevance to the research question. Search strategies used for each database were documented in Appendix 2, Section 10.2.4 of the MS.

Searches were limited to studies published between 1970 and May 2013, with the justification that the earliest melanoma trial was published in 1972. In view of the therapies eligible for inclusion in the review, this was appropriate. The searches were not limited by language.

In Appendix 6 (Section 10.6), under 'Search strategy for Section 6.8 (Non-RCT evidence)', the MS states "Not applicable" for each of the subsections; however, non-RCT evidence was included in the review. Therefore, it is unclear in the MS how these studies were identified. The ERG asked the manufacturer to clarify how non-RCT evidence was identified; the manufacturer responded that non-RCT evidence was not searched for systematically because direct RCT evidence was identified for comparator therapies, therefore, non-RCT evidence was only required for ipilimumab 3 mg/kg monotherapy. The manufacturer stated that the non-RCT evidence is derived from ongoing trials or pooled data analysis that are not currently available in full, in the public domain.

Generally, the search strategies were well reported in the MS. The exact strategies used along with the databases searched, the database service provider, dates of searches and date span of the searches are all clearly reported. Searches for both published and unpublished literature were carried out and all of the NICE specified databases were searched.

The textword and subject headings used in the search strategies for ipilimumab and the comparator drugs were appropriate. Various synonyms and spelling variations for melanoma or skin cancer have been included in the strategy. However the terms neoplas$ and tumor$ have not been included. Boolean operators, truncation and field searches have been used appropriately.

The search strategies were limited to RCTs or systematic reviews using study design filters developed by SIGN. The SIGN filters have been developed in-house and have not been independently validated. Other validated search filters with high sensitivity are freely available and could have been used. A further issue with restricting the clinical effectiveness searches to RCTs or systematic reviews is that adverse events data from other study types may not have been identified.

Unpublished literature was sought from Clinicaltrials.gov using an appropriate search strategy. In addition, several conference proceedings were searched; however details of how they were searched were not reported.

Although minor issues have been raised with the search strategies for clinical effectiveness, it is unlikely that any relevant RCTs have been missed. Non-RCT evidence was not searched for systematically; however, the manufacturer stated that non-RCT evidence for ipilimumab 3 mg/kg monotherapy is derived from ongoing trials or pooled data analysis that are not currently available in full, in the public domain.

The inclusion criteria relating to the population of interest appear to have been appropriate and in line with both the decision problem and the marketing authorisation of ipilimumab. The inclusion criteria relating to outcomes of interest appear to have been appropriate, and included all those specified in the scope.

As discussed in Section 3.3, it is unclear why dabrafenib was added to the inclusion criteria relating to interventions of interest, as this was not presented in Table 4 of the MS, which presented the decision problem addressed in the submission. The manufacturer stated that it was included as a clinical comparator, as it is licensed for monotherapy for the treatment of adult patients with unresectable or metastatic melanoma with a BRAF V600 mutation, and is currently scheduled for a NICE technology appraisal. However, it does not meet the reference case for comparators, defined as "therapies routinely used in the NHS, including technologies regarded as current best practice".

Prospective non-RCTs and observational studies were listed under 'exclusion criteria' for the review of the clinical effectiveness data; however, two ongoing retrospective observational studies and a pooled analysis of a subgroup of patients included in four RCTs were included in the MS, due to the lack of RCT evidence directly investigating ipilimumab 3 mg/kg monotherapy in untreated patients. Non-RCT evidence was not searched for systematically (see Section 4.1.1).

A flow chart of the study selection process was presented in the MS (Figure 5), which stated that six original RCTs were included in a quantitative synthesis. An additional record of an ongoing trial was included in the qualitative synthesis, along with 9 kin publications, related to the six included RCTs. The complete list of relevant RCTs was as follows:

- CA184-024[5]
- MDX010-08[6]
- CA184-004[13]
- CA184-022[12]
- CA184-169[17]
- BREAK-3[8]

4.1.5 Evidence synthesis

The manufacturer described the results of the individual trials separately, which was appropriate in view of the differences in study design, participant and intervention characteristics.

A mixed treatment comparison of ipilimumab (10 mg/kg plus DTIC), and the BRAF inhibitors vemurafenib and dabrafenib was presented, using the CA184-024, BRIM-3 and BREAK-3 trials, with DTIC being the common comparator treatment in each of the trials.[5, 7, 8] Indirect pairwise comparisons, determined using the Bucher method, were also presented. The ERG does not consider these analyses to be appropriate because of differences in patients' baseline characteristics between the ipilimumab, vemurafenib and dabrafenib trials, resulting in patients in the vemurafenib and dabrafenib trials having a worse prognosis than patients in the ipilimumab trial (differences in the proportion of patients with elevated serum lactate dehydrogenase levels and M1c disease).

4.1.6 Conclusions from critique of systematic review methods

The search strategy for RCT evidence was appropriate, comprehensive and well documented. The inclusion criteria for the population and outcomes of interest were appropriate, but it is not clear why dabrafenib was included as an intervention of interest, since it was not included in the decision problem and is not currently routinely used in the NHS. The evidence on dabrafenib does not inform the clinical effectiveness of ipilimumab or vemurafenib and the comparison between ipilimumab and dabrafenib is not used in the economic model. The methods used for study selection were not reported; therefore it is unclear whether they were susceptible to error or bias. Two of the studies which were stated as being included in the quantitative synthesis were excluded from the review due to the low patient numbers in each treatment arm that met the criteria for inclusion in the review. No relevant RCTs of ipilimumab and comparator therapies for the treatment of adult patients with previously untreated advanced malignant melanoma appear to have been overlooked. Adequate details of the included studies were presented and quality assessment was appropriate. Additional evidence was presented in the MS from trials that were excluded from the review; CA184-004, CA184-022 and MDX010-020.[12, 13, 18]

monotherapy treatment arms of four RCTs (MDX010-08, CA184-004, CA184-022 and MDX010-020).[11] The MDX010-08 trial was included in the MS separately, in addition to being included in the pooled dataset.

The study design and participant baseline characteristics for the three included non-RCT studies are presented in Table 4.4 (Tables 22 and 24 of the MS). The dosing schedule of ipilimumab was the same in the three non-RCTs and reflected the license. There were some clinically significant differences in the baseline characteristics of patients between the three datasets; including differences in performance status, disease stage, presence of brain metastases, duration of melanoma and prior immunotherapy, as shown in Table 4.4. Each of the non-RCTs assessed overall survival as an outcome.

The manufacturer compared the results of these uncontrolled studies with the DTIC arm of the CA184-024 trial, which is inappropriate, because of the differences in study design and patient characteristics.

pathological diagnosis of melanoma to first dose of study therapy was twice as long in the ipilimumab monotherapy group), follow-up was adequate, an appropriate intention-to-treat analysis was performed and there were no unexpected imbalances in drop-outs between treatment groups.

The BREAK-3 trial was a reasonably large open-label RCT; randomisation appears to have been carried out appropriately, the treatment groups were generally similar at baseline, an appropriate intention-to-treat analysis was performed and there were no unexpected imbalances in drop-outs between treatment groups.[8] However, median follow-up was only five months in the dabrafenib group.

The BRIM-3 trial was a large open-label RCT in which an independent review committee was blinded to treatment allocation; it is unclear whether randomisation was carried out appropriately, the treatment groups were generally similar at baseline, an appropriate intention-to-treat analysis was performed and there were no unexpected imbalances in drop-outs between treatment groups.[7] However, median follow-up was less than a year in both groups.

The MS also referred to trials CA184-004 and CA184-022, despite these full trials not meeting the inclusion criteria for the review (they included only a small number of chemotherapy-naïve patients). The CA184-004 trial was a small double-blind phase II trial, underpowered to detect a statistically significant difference in overall survival; it is unclear whether randomisation was carried out appropriately, the treatment groups were generally similar at baseline (although more patients in the 10 mg/kg treatment group had stage M0 or M1c disease), median follow-up was less than 9 months in both groups.[13] The CA184-022 trial was a reasonably large double-blind randomised dose-ranging phase II trial; randomisation appears to have been carried out appropriately, the treatment groups were generally similar at baseline (although more patients in the 10 mg/kg treatment group had elevated levels of lactate dehydrogenase and more patients in the 0.3 mg/kg treatment group had stage M1c disease), an appropriate intention-to-treat analysis was performed and there were no unexpected imbalances in drop-outs between treatment groups.[12]

4.2.2.2 Non-RCT evidence

Results of the quality assessment for the non-RCT evidence were presented in Appendix 7 of the MS. Whilst the quality of the three studies was generally adequate, there is a high risk of bias associated with these study designs, relative to an RCT. The manufacturer acknowledged the potential biases introduced through the use of retrospective observational data and pooled analyses. Due to the nature of the evidence and the inclusion of data from the MDX010-08 trial in the pooled analysis, in addition to being presented as part of the RCT evidence

, the ERG does not consider that these studies provide additional relevant reliable data.

In the manufacturer's response to the ERG's points for clarification document they also presented the results of an additional pooled analysis.[14] However, no methods for the pooled analysis were presented; therefore, it is not possible to determine the reliability of the results of this pooled analysis.

4.2.3 Summary of the results of the included trials

4.2.3.1 RCT evidence

Overall survival results for patients enrolled in the four RCTs are presented in Table 4.5 (Table 16 of the MS). The individual study results are then presented separately.

Table 4.5 Overall survival of patients enrolled in CA184-024, MDX010-08, BREAK-3 and BRIM-3

Trial no. (acronym)	Treatment arm	Median OS, months (95% CI)	Hazard Ratio (95% CI)
CA184-024[5]	Ipilimumab 10mg/kg + DTIC DTIC	11.2 (9.4, 13.6) 9.1 (7.8, 10.5)	0.72 (0.59, 0.87)
MDX010-08[6]	Ipilimumab 3mg/kg + DTIC Ipilimumab 3mg/kg	14.3 (10.2, 18.8) 11.4 (6.1, 15.6)	0.75 (0.45, 1.24)
BREAK-3[8]	Dabrafenib DTIC	Not reported	0.61 (0.25, 1.48)
BRIM-3[7]	Vemurafenib DTIC	13.2 (12.0, 15.0) 9.6 (7.9, 11.8)	0.62 (0.49, 0.77)

Ipilimumab 10 mg/kg plus DTIC versus DTIC plus placebo (CA184-024)

The CA184-024 trial demonstrated that ipilimumab 10 mg/kg plus DTIC (including a maintenance phase of ipilimumab therapy) significantly increased overall survival, compared with DTIC plus placebo, in patients with previously untreated, advanced melanoma.[5] Median overall survival was increased by 2.1 months (11.2 months versus 9.1 months; $p<0.001$), as shown in Table 4.5. Overall survival rates were higher with ipilimumab plus DTIC than DTIC plus placebo at one year (47.6% versus 36.4%), two years (28.9% versus 17.8%), three years (21.3% versus 12.1%), four years (19.1% versus 9.7%) and five years (18.2% versus 8.8%).[28] The Kaplan-Meier curve for overall survival is presented as Figure 4.1 (Figure 10 of the MS).

The ERG considers a conventional three-state model structure for progression-free survival, post-progression and death used in most cancers to be more appropriate as it avoids the use of arbitrary assumptions about second- and third-line treatment. The implications of the manufacturer's approach on the cost-effectiveness results are presented and discussed in Section 6. The assumptions made with regard to the treatment sequencing and the ERG's concerns around these are discussed in detail in the following sections.

5.2.5 Treatment effectiveness and extrapolation

The transitions between lines of treatment and the proportion of patients who die in each cycle were informed by extrapolated clinical trial data. Parametric survival curves were fitted to empirical data on overall survival (OS) and progression-free survival (PFS) to enable the observed data to be extrapolated beyond the follow-up period of the trial in order to estimate mean survival times required for the cost-effectiveness analysis. For ipilimumab and DTIC the main source of data used to inform OS and PFS was the respective treatment arms of the CA184-024 trial, which investigated the efficacy of ipilimumab 10mg/kg in addition to DTIC therapy (used to inform the efficacy of ipilimumab) compared with DTIC monotherapy (used to inform the efficacy of DTIC). For vemurafenib, separate, independent OS and PFS data were used to model the comparative efficacy of ipilimumab and vemurafenib, which was informed by the BRIM-3 trial which investigated the efficacy of vemurafenib compared with DTIC in patients testing positive for the BRAF V600 mutation. Clinical disease progression using PFS data was used as a proxy for determining change in therapy, i.e. the movement to the next line of treatment. For second-line treatments, the OS was based on the first-line survival curves but adjusted downwards to account for poorer outcomes on second- line therapy using a constant proportional hazard derived from the expected life years of second-line ipilumumab from TA268. The duration of response to second-line treatments was based on the number of pre-progression life years for second-line ipilumumab from TA268.[1] Third line treatment was assumed to be best supportive care (BSC), which consisted of a proportion of patients on 'no treatment' and a proportion on commonly prescribed chemotherapy drugs, including DTIC. The OS for patients who survived to receive third-line BSC was assumed to be the same as those on first-line treatment who had not progressed to next line of therapy.

The ERG has identified several specific issues which relate to the subsequent use of these data in the economic model. The key issues identified by the ERG include:

- The appropriateness of assuming that the absence of comparative efficacy evidence for ipilumumab 3mg/kg in previously untreated advanced melanoma implies equal efficacy to ipilimumab 10mg/kg;

effectiveness. The ERG has some concerns that no direct EQ-5D data were used in the model and that the algorithm by Rowen et al may not be sufficiently generalisable to this patient population as it is based upon a dataset from an international study of patients with multiple myeloma with the valuation study being carried out in the UK.

The utility values used in the MS were said to be consistent with the utilities used in the second-line submission of ipilimumab (TA268);[1] however the ERG notes that the base-case utility values in TA268 were based on progression and not on proximity to death. The manufacturer's rationale for using time-dependent utilities was based on the analysis of patient-level HRQoL data from the CA184-024 trial, which showed a strong correlation between HRQoL and time-to-death.[5] In addition, the approach allowed a utility decrement for active treatment with ipilimumab to be incorporated within the analysis. This decrement was assumed to represent the impact of treatment-related adverse events on HRQoL. In the absence of evidence for vemurafenib, the manufacturer assumed that the same decrement could be applied to vemurafenib on the basis that the number of patients experiencing grade 3+ adverse events was similar. The manufacturer also assumed that the utility values at each time point before death for vemurafenib were identical to those of ipilimumab. To assess the validity of this assumption, the ERG requested additional clarification from the manufacturer. In response, the manufacturer stated that this assumption was necessary as (i) no data were available for vemurafenib; (ii) the number of serious adverse events (SAEs) for patients treated with ipilimumab and vemurafenib were not significantly different; and (iii) the long-term impact of adverse events is likely to be greater with vemurafenib as it is continuously dosed whereas ipilimumab is not.[14]

The utility weights used in each sub-health state in the manufacturer's model are summarised in Table 5.4. The utility estimates were applied to the proportion of patients within each sub-health state during each cycle of the model, in order to estimate QALYs for each treatment strategy.

Table 5.10 Base case results, CA184-024 data for ipilimumab – BRAF V600 mutation-positive

Incremental pair-wise comparison with ipilimumab

Technologies	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	ICER vs. baseline (A)	Incremental analysis
DTIC	£44,267	1.56				
Ipilimumab	£68,033	2.31	£23,766	0.75	£31,559	£31,559
Vemurafenib	£80,658	2.13	£12,625	-0.18	£63,534	Dominated

Table 5.11 Base case results, CA184-024 data for ipilimumab – BRAF V600 mutation-negative

Technologies	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental QALYs	ICER (£) Ipi vs. comparator
Ipilimumab	£57,760	3.3522	2.3527			
DTIC	£44,267	2.1476	1.5570	£13,492	0.7957	£16,958

5.2.8.2 Sensitivity analyses

Deterministic sensitivity analysis

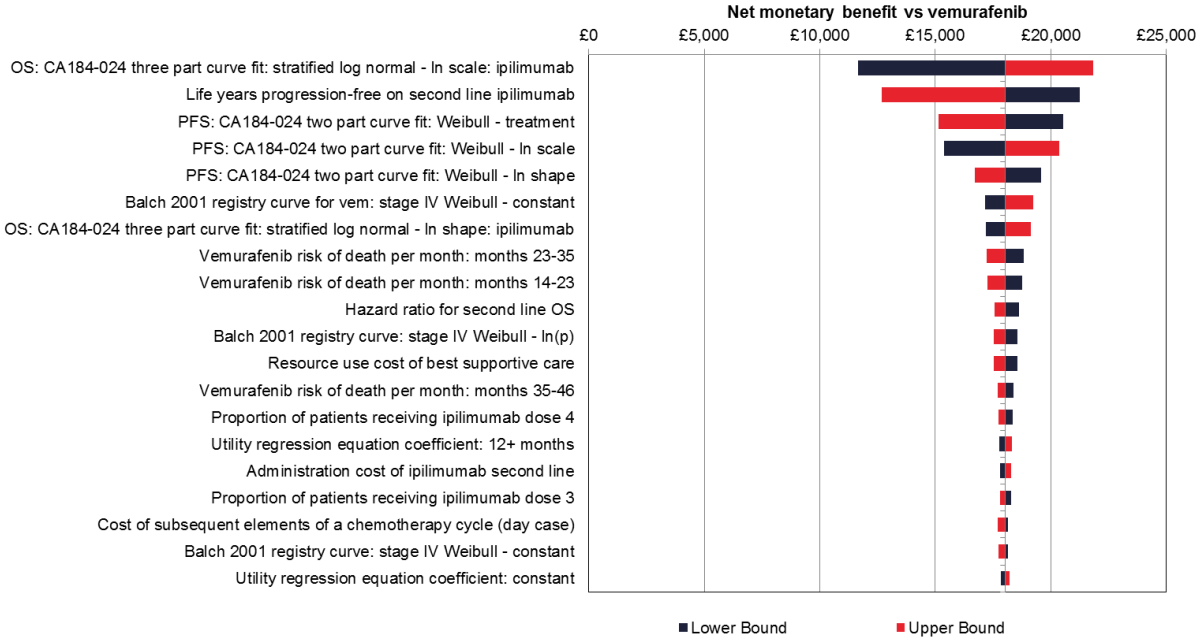
The manufacturer presented a series of one-way deterministic sensitivity analyses (DSA) to assess the impact of uncertainty on key input parameters on the ICERs. The manufacturer only presented sensitivity analyses for the BRAF V600 mutation-positive population since the ICER for ipilumumab compared with DTIC in the BRAF V600 mutation-negative population was lower than in the BRAF V600 mutation-positive population; therefore the sensitivity analyses for the mutation-positive population was assumed to represent a worst case scenario.

Figure 5.14 shows a tornado diagram of the 20 parameters which the manufacturer deemed to be the most influential parameters on the assessment of cost-effectiveness for ipilimumab compared with vemurafenib. The parameters were varied between their upper and lower bounds and the effect on incremental net benefit was presented. The manufacturer used two different willingness to pay thresholds for the calculation of incremental net benefit; £30,000 per QALY for the comparison with vemurafenib and £50,000 per QALY for the comparison with DTIC. The ERG notes that the use of the higher threshold for the comparison with DTIC will make ipilimumab appear to have a higher probability of being cost-effective.

According to the manufacturer's DSA, the model proved to be most sensitive to the parameters used to model overall survival of ipilimumab and DTIC, the time spent on second-line treatment and the time spent on first-line treatment with ipilimumab relative to DTIC.

Figure 5.14 Tornado diagram containing 20 most influential parameters vs. Vemurafenib

CA184-024 data for ipilimumab



Probabilistic sensitivity analysis

The manufacturer performed probabilistic sensitivity analysis using 1,000 simulations. Scatter plots and cost-effectiveness acceptability curves were presented in the MS. The manufacturer's PSA concluded that the probability of ipilimumab being the most cost-effective treatment compared with DTIC was 96% at a willingness to pay threshold of £50,000. At a threshold of £30,000 per QALY, the probability that ipilimumab is cost-effective is 40%.

The mean ICERs that were derived from the probabilistic sensitivity analysis were similar to those of the deterministic analysis as shown in Table 5.12.

Table 5.12 Results of PSA (mean over 1,000 runs) (Table 78, P199 in the MS)

Ipi data source	Ipilimumab		Vemurafenib		DTIC		ICER vs. vem	ICER vs. DTIC
	Cost	QALYs	Cost	QALYs	Cost	QALYs		
CA184-024	£67,938	2.29	£80,629	2.12	£44,298	1.54	Dominated	£31,619
Pooled chemo-therapy-naïve	£68,736	2.38	£78,156	2.08	£42,898	1.54	Dominated	£30,756

5.2.8.3 Scenario analyses

The MS also included a series of scenario analyses that were performed to check the robustness of the model to uncertainty related to structural assumptions, e.g. time horizon, treatment pathway, approach to modelling drug dosing, utilities and treatment efficacy assumptions. The main analyses are reported

be similar to the values used in the economic model for the appraisal of second-line ipilimumab (TA268),[1] which were validated by clinicians. However, the ERG notes that no scenarios were presented by the manufacturer to demonstrate this consistency and the approach taken to model HRQoL differs from that used in the base-case analysis of TA268[1] by using time-dependent instead of progression-based utility values. The implications on the cost-effectiveness results have not been explored by the manufacturer.

Furthermore, the manufacturer has not validated the use of independent treatment arms from different clinical trials (CA184-024[5] for ipilimumab and DTIC, and BRIM-3[7] for vemurafenib) for the assessment of cost-effectiveness of ipilimumab compared with vemurafenib. The generalisability of the CA184-024 clinical study to the UK population has only been briefly discussed by the manufacturer, which states that it is generally comparable to the UK population receiving first- line therapy as reported in Lorigan et al (2013),[43] but that the trial population may be slightly more severe.

5.3 Summary of uncertainties and issues from the cost-effectiveness analysis

The ERG considered the manufacturer's economic submission to meet the requirements of the NICE reference case. However, the ERG does not consider the economic model structure appropriate to address the decision problem in light of the lack of evidence available on the sequencing of treatments for previously untreated malignant melanoma. Although the structure of the model is not wholly appropriate for the evidence base that is available for first-line ipilimumab, it is possible to 'turn off' the treatment sequencing part of the manufacturer's model, which results in a direct comparison of first-line ipilimumab with DTIC or vemurafenib in terms of OS and PFS in previously untreated patients.

The main concerns expressed by the ERG relate to the following issues:

1. *The treatment sequencing approach*

Although conceptually it is valid to consider the sequencing of treatments and the ERG recognises that the approval of multiple treatments for advanced melanoma may lead to a sequential use of them in clinical practice. However, there are no reported results from any clinical trials comparing the sequential use and relative effect of treatments used in sequence for previously untreated advanced melanoma. This lack of evidence has led the manufacturer to oversimplify the treatment sequencing approach by using arbitrary assumptions about overall survival on second and third line treatments. In particular, the manufacturer's approach presents a switch in OS curves between first-, second-, and third-line (BSC) treatments, which is inconsistent and does not appear to have a clinical rationale underpinning the choice of curves in each line.

6.3 Exploring the robustness of the cost-effectiveness results to specific assumptions and additional uncertainties identified by the ERG

During the critique of the MS in Section 5.2.5, the ERG identified a number of uncertainties in relation to the comparative effectiveness estimates applied in the model. In particular, the ERG noted that the evidence for the cost-effectiveness of ipilimumub within it licensed dose of 3mg/kg every 3 weeks for a total of 4 doses was based on the efficacy of ipilimumab 10mg/kg in addition to DTIC therapy from the CA184-024 trial,[5] and the use of separate, independent arms from different clinical trials (i.e. breaking randomisation) was used to establish the relative efficacy with vemurafenib. Furthermore, the method used for the comparison with vemurafenib was inconsistent with the values reported from the indirect comparison analysis, which was also undertaken by the manufacturer. Given the importance of the clinical evidence as a driver of cost-effectiveness, the ERG has examined a set of individual comparisons for both the BRAF V600 mutation-negative and –positive populations where the assumptions on the clinical evidence have been varied. The ERG also noted in Section 5.2.7.3 that the costs of ipilimumab per weekly cycle were considerably higher with ipilimumab treatment second-line compared with ipilimumab first-line. The implications of this cost difference on the cost-effectiveness results are explored below. The ERG has also explored in a scenario analysis the implications of using time-to-death HRQoL outcomes by alternatively assigning utility values to pre- and post-progression health states.

6.3.1 Comparison of overall survival with ipilimumab 3mg/kg and ipilimumab 10mg/kg

In the absence of sufficient comparative efficacy evidence for previously untreated patients, the manufacturer assumed equivalent efficacy between 3mg/kg dose of ipilimumab and 10mg/kg dose in terms of both overall and progression-free survival. However, as discussed in Sections 4.2.3.1 and 4.5, from the CA184-022 trial of previously treated and untreated patients, median overall survival appeared to be better in the 10 mg/kg treatment group but the difference was not statistically significant.[12] A very small trial of previously treated and untreated patients (CA184-004) found no meaningful differences in clinical activity between the two ipilimumab doses.[13] The ERG requested further clarification on this point and asked the manufacturer to provide a comparison of ipilimumab 3mg/kg and ipilimumab 10mg/kg in both the treatment-naïve subgroup population and the whole population (both previously treated and untreated) in all clinical trials.[14] In response the manufacturer presented results from a pooled data analysis comparing OS profiles of ipilimumab 3mg/kg (MDX010-20; CA184-022) and 10mg/kg (CA184-007; CA184-008; CA184-022) for a mixed (both previously treated and untreated) patient population. The Kaplan-Meier curves for this pooled data analysis are presented in Figure 4.6, Section 4.2.3.2 (a digitisation of the curves is also presented below in Figure 6.8). The manufacturer reported that no significant difference in survival was observed between the ipilimumab 3mg/kg and 10mg/kg treatment arms in the whole population