CONFIDENTIAL UNTIL PUBLISHED

Evidence Review Group Report commissioned by the NIHR HTA Programme on behalf of NICE

Nivolumab for treating advanced (unresectable or metastatic) melanoma

Produced by Southampton Health Technology Assessments Centre (SHTAC)

Authors Keith Cooper, Senior Research Fellow, SHTAC
Neelam Kalita, Research Fellow, SHTAC
Elke Streit, Research Fellow, SHTAC
Jonathan Shepherd, Principal Research Fellow, SHTAC
Maria Chorozoglou, Senior Research Fellow, SHTAC
Geoff Frampton, Senior Research Fellow, SHTAC

Correspondence to Jonathan Shepherd
Southampton Health Technology Assessments Centre (SHTAC)
University of Southampton
First Floor, Epsilon House
Enterprise Road, Southampton Science Park
Southampton
SO16 7NS
UK
www.southampton.ac.uk/shtac

Date completed 28th October 2015

Source of funding: This report was commissioned by the NIHR HTA Programme as project number 15/06/11.

Declared competing interests of the authors

None

Acknowledgements

We would like to thank Dr Pippa Corrie (Consultant and Associate Lecturer in Medical Oncology, Addenbrooke's Hospital, Cambridge, UK) for providing clinical advice during the production of this report, and for commenting on a draft of the report.

We also thank: Karen Welch, Information Scientist, SHTAC, for appraising the literature search strategies in the company's submission, running an update of the company's searches and searching for ongoing studies; Dr Jo Picot, Senior Research Fellow, SHTAC, for assistance with appraising the company submission; Professor Joanne Lord, Director, SHTAC, for acting as internal editor for the draft report; and Dr Vicky Copley, independent statistical adviser, for providing statistical advice to the team.

Rider on responsibility for report

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

This report should be referenced as follows:

Cooper K, Kalita N, Streit E, Shepherd J, Chorozoglou M, Frampton G. Nivolumab for treating advanced (unresectable or metastatic) melanoma: A Single Technology Appraisal. Southampton Health Technology Assessments Centre (SHTAC), 2015.

Contributions of authors

Keith Cooper (Senior Research Fellow) critically appraised the health economic systematic review, critically appraised the economic evaluation and drafted the report. Neelam Kalita (Research Fellow) critically appraised the health economic systematic review, critically appraised the economic evaluation and drafted the report. Elke Streit (Research Fellow) critically appraised the clinical effectiveness systematic review and drafted the report. Jonathan Shepherd (Principal Research Fellow) critically appraised the clinical effectiveness systematic review, drafted the report, project managed the review and is the project guarantor. Maria Chorozoglou (Senior Research Fellow) critically appraised the health economic systematic review, critically appraised the economic evaluation and drafted the

report. Geoff Frampton (Senior Research Fellow) critically appraised the clinical effectiveness systematic review and drafted the report.

Word count:

38,822

[redacted]

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AE	Adverse events
AIC	Akaike information criterion
AJCC	American Joint Committee on Cancer
AUC	Area under the curve
BIC	Bayesian information criterion
CHMP	Committee for Medical Products for Human Use
CI	Confidence interval
CONSORT	Consolidated Standards of Reporting Trials
CR	Complete response
CS	Company's submission
CSR	Clinical study report
DC	Discontinuation
DIC	Deviance Information Criterion
DTIC	Dacarbazine
CTLA-4	Cytotoxic T-lymphocyte-associated protein 4
ECOG	Eastern Cooperative Oncology Group
EORTC-QLQ-C30	European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire-Core 30
EPAR	European Public Assessment Report
EQ-5D	EuroQoL five dimension questionnaire
ERG	Evidence Review Group
HRQoL	Health-related quality of life
HR	Hazard ratio
ICC	Investigator's choice of chemotherapy
ICER	Incremental cost-effectiveness ratio
IRRC	Independent radiological review committee
ITT	Intention-to-treat
LDH	Lactate dehydrogenase
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
OR	Odds ratio
ORR	Objective response rate
OS	Overall survival
PAS	Patient Access Scheme
PD-L1	Programmed death receptor ligand 1
PFS	Progression-free survival
PP	Per protocol
PR	Partial response
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PS	Performance status
PSA	Probabilistic sensitivity analysis
PSS	Personal Social Services
QALY	Quality-adjusted life year
RCT	Randomised controlled trial
RECIST	Response Evaluation Criteria in Solid Tumors
SAE	Serious adverse event
SD	Standard deviation
SE	Standard error
STA	Single Technology Appraisal
SmPC	Summary of product characteristics

TA	Technology Appraisal
TRAE	Treatment-related adverse event
TRSAE	Treatment-related serious adverse event

SUMMARY

Scope of the company submission

The company's submission (CS) reflects the scope of the appraisal issued by the National Institute for Health and Care Excellence (NICE). The submission assesses the clinical effectiveness and cost effectiveness of nivolumab monotherapy compared to BRAF inhibitors (dabrafenib and vemurafenib for BRAF V600 mutation-positive melanoma), ipilimumab, dacarbazine (DTIC), and to best supportive care for the treatment of adults with advanced (unresectable or metastatic) melanoma.

Summary of submitted clinical effectiveness evidence

The company's systematic review of clinical effectiveness identified three relevant phase III RCTs of nivolumab monotherapy. In these, nivolumab was administered by intravenous infusion at a dosage of 3mg/kg every two weeks.

- The CheckMate 066 trial compared nivolumab with 1000mg/m^2 DTIC, administered every three weeks by intravenous infusion. Participants were treatment-naïve patients who did not have a BRAF mutation.
- The CheckMate 067 trial compared nivolumab with 3mg/kg ipilimumab, administered every three weeks by intravenous infusion. Participants were treatment-naïve patients, and BRAF mutation-negative as well as BRAF mutation-positive patients were enrolled in this trial. Nivolumab in combination with ipilimumab was also investigated in this study, but is outside the NICE scope and therefore not included in the CS.
- The CheckMate 037 trial was an open-label study that compared nivolumab with the investigator's choice of chemotherapy (ICC), either 1000mg/m^2 DTIC every three weeks or paclitaxel 175mg/m^2 combined with carboplatin area under the curve 6 every three weeks. Participants were patients who progressed on or after prior ipilimumab, or ipilimumab and a BRAF inhibitor if they were BRAF mutation-positive.

The primary outcome in all three studies was overall survival (OS). Additional primary outcomes were progression-free survival (CheckMate 067) and objective response rate (CheckMate 037). The trials were judged by the Evidence Review Group (ERG) to be of generally good methodological quality. The ERG believes that it is likely that the company has identified all relevant RCTs.

The CS reports the effects of nivolumab across a range of outcomes relevant to the NICE scope and decision problem, summarised below. All CheckMate trials are still ongoing for extended follow-up in order to generate evidence on longer-term outcomes, including OS, PFS, and health-related quality of life (HRQoL).

OS data are available for the CheckMate 066 trial. There was a significant reduction in all-cause mortality with nivolumab when compared to DTIC. At the time of database lock (August 2014), the median OS had not been reached, i.e. more than half of the nivolumab-treated patients were still alive, whereas most of the patients treated with DTIC had already died (median OS = 10.84 months).

Significant PFS benefit was observed when nivolumab was compared with DTIC (CheckMate 066) or ipilimumab (CheckMate 067), but no difference in PFS was detected between nivolumab and ICC in CheckMate 037, presumably due to the immaturity of the PFS data in the latter trial.

In terms of ORR, there was significant benefit of nivolumab over comparator drugs in all three CheckMate trials. More patients treated with nivolumab experienced complete response than those treated with alternative drugs, although the total number of patients with complete response was low in all study groups (<10%). Furthermore, treatment response was found to be more durable in nivolumab-treated patients compared to patients treated with DTIC, ipilimumab, or ICC, with the longest duration of response observed in the CheckMate 067 nivolumab group exceeding 12 months at the time of reporting.

There was also a significant change in tumour burden in nivolumab-treated patients. More patients in the nivolumab groups of the CheckMate trials experienced reductions in tumour size and achieved at least a partial response, compared with patients treated with DTIC, ipilimumab, or ICC.

In all three CheckMate trials, patients were able to continue treatment beyond progression if experiencing clinical benefit, and a proportion of those treated with nivolumab continued to respond to the drug (ORR up to 27.0% in CheckMate 037).

Interim analyses of health-related quality of life (HRQoL) were available only for the CheckMate 066 trial. Patients receiving nivolumab tended to have higher HRQoL scores at baseline than those receiving DTIC but the statistical significance of the difference is questionable. Although nivolumab appeared to improve some aspects of HRQoL relative to

baseline scores when assessed on the EQ-5D and different subscales of the EORTC-QLQ-C30 questionnaire (European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire-Core 30 questionnaire), there is no consistent evidence that nivolumab had a sustained effect on HRQoL. The company concluded that nivolumab does not impair HRQoL (relative to baseline), and the ERG agrees that this is a reasonable conclusion based on the interim data that are available.

Pre-defined subgroup analyses were undertaken for most baseline characteristics, and outcomes were in favour of nivolumab for most subgroups. Nivolumab-treated patients experienced benefit regardless of programmed death receptor ligand 1 (PD-L1) status compared to patients treated with the comparator drugs, although benefit was highest in PD-L1-positive patients, with lower mortality rates and longer PFS compared to patients with PD-L1-negative status. Subgroup analyses by BRAF mutation demonstrated a benefit for nivolumab compared to ipilimumab in terms of PFS and ORR in CheckMate 067, regardless of BRAF mutation status, but patients in the BRAF mutation-negative group experienced higher benefit than those with BRAF positive status.

The proportion of patients who experienced adverse events (AEs) was generally similar between nivolumab and the comparator drugs. Nearly all patients experienced at least one AE of any grade, regardless of treatment allocation, and the majority of AEs were treatment-related. Higher grade and serious AEs occurred less frequently in nivolumab-treated patients, and a smaller proportion of patients discontinued nivolumab treatment due to treatment-related AEs (TRAEs) compared to patients treated with any of the comparator drugs. The most frequently reported TRAEs among nivolumab-treated patients were fatigue, pruritus, rash, diarrhoea, and nausea. Treatment-related serious AEs (TRSAEs) included hyperglycaemia, vomiting, pyrexia, and pneumonitis. These were not reported by more than two patients in any CheckMate trial, and most were resolved. One death due related to nivolumab treatment was reported.

Indirect comparisons were conducted using selected RCTs from the company's systematic review of clinical effectiveness. Two separate evidence networks were created, for the comparison with ipilimumab and palliative chemotherapy (BRAF mutation-negative patients), and for the comparison with BRAF inhibitors (vemurafenib and dabrafenib, BRAF mutation-positive patients). The networks used patient-level data / 'pseudo' patient-level data (BRAF mutation-negative / mutation-positive patients, respectively) from the trials to inform covariate-adjusted parametric survival models used directly in the economic model.

Summary of submitted cost effectiveness evidence

The CS includes:

i) A review of published economic evaluations of nivolumab for advanced melanoma

ii) An economic evaluation undertaken for the NICE STA process. The cost effectiveness of nivolumab is compared to that ipilimumab and DTIC for BRAF mutation-negative patients and compared to ipilimumab, dabrafenib and vemurafenib for BRAF mutation-positive patients.

A systematic search of the literature was conducted by the company to identify economic evaluations of nivolumab for advanced melanoma. The review did not identify any relevant studies.

The cost effectiveness analysis (CEA) uses a semi-Markov model to estimate the cost-effectiveness of nivolumab compared with DTIC and ipilimumab for BRAF mutation-negative patients and with dabrafenib, ipilimumab and vemurafenib for BRAF-mutation-positive patients with advance melanoma. The model adopted a lifetime horizon of 40 years and a cycle length of one week. The model consisted of three health states: pre-progression, progression and death.

The economic evaluation used data from the CheckMate 066 trial. The company conducted covariate-adjusted indirect comparisons between comparators using patient-level data. These data were used to estimate time to progression (TTP), post-progression survival (PPS) and pre-progression survival (PrePS), which were used to derive the transition probabilities between health states.

Results of the economic model were presented as incremental cost per quality-adjusted life years (QALY) and incremental cost per life years gained. Three of the comparators (ipilimumab, dabrafenib and vemurafenib) have a confidential patient access (PAS) scheme. Results were presented at the list price and at the estimated PAS prices. The results of the cost effectiveness analysis for BRAF mutation-negative patients at the list price showed that nivolumab is cost effective compared to ipilimumab and DTIC and for BRAF mutation-positive patients nivolumab is cost effective compared to dabrafenib, vemurafenib and ipilimumab at a willingness to pay threshold of £30,000 per QALY.

The company performed a range of deterministic and probabilistic sensitivity analyses to assess model uncertainty. The base case results were robust to uncertainties in the key model parameters and assumptions, except for changes in the maximum treatment duration for nivolumab. The PSA showed that there is 87% and 99% probability of nivolumab being cost-effective for BRAF-mutation-negative patients at a willingness to pay threshold of £30,000 and £50,000 per QALY gained, and a 100% probability of nivolumab being cost effective for BRAF-mutation-positive patients for both thresholds.

Commentary on the robustness of submitted evidence

Strengths

The decision problem in the company submission generally accords with the NICE scope. However, the ERG notes that the economic analysis includes DTIC as a comparator in the BRAF mutation-negative analysis, but not in the BRAF mutation-positive analysis, with no apparent justification.

The company's systematic review of clinical-effectiveness followed standard procedures and is of good quality. The ERG is not aware of any additional relevant published trials that could be included.

The three key CheckMate RCTs were well-designed and well-conducted and provide an appropriate evidence base to inform the assessment of clinical-effectiveness and cost-effectiveness of nivolumab.

The structure of the economic model was appropriate, comprehensive and reflected the clinical pathway for patients with advanced melanoma. The model was well-structured and provided the relevant data sources in a transparent way.

The methods chosen for the analysis were generally appropriate and conformed to NICE methodological guidelines.

The company performed a wide range of sensitivity analyses including one-way, probabilistic and scenario analyses to assess model uncertainty.

Weaknesses and areas of uncertainty

All three of the key RCTs included by the company in their systematic review of clinical-effectiveness are ongoing with further follow-up results expected to be published in the next year. Consequently, some of the results reported in the CS are from interim time points, in some cases based on relatively small numbers of patients or events, and are considered to be relatively immature due to lack of follow-up. This is notably for overall survival, one of the key outcomes that informs the assessment of cost-effectiveness in the CS.

The comparative efficacy of nivolumab with the comparator treatments in the NICE scope is uncertain due to a lack of available head-to-head data from clinical trials. The company's indirect comparison is complex and is based upon a number of assumptions and survival data extrapolations. Some of these assumptions appear reasonable and are noted by the CS to have been accepted in previous NICE appraisals of treatments for advanced melanoma. However, there is some uncertainty regarding the assumption and that there is no difference in treatment effects for nivolumab by BRAF mutation status. This is of significance as evidence from the CheckMate 066 trial, which included BRAF mutation-negative patients, was indirectly compared with evidence from a BRAF inhibitor trial, by definition including BRAF mutation-positive patients, and informed cost-effectiveness estimates for the BRAF mutation-positive patient group.

There is some uncertainty about the survival curves that best represent long-term overall survival and progression free survival, due to the short follow-up data currently available for the CheckMate trials.

The time spent on treatment is a key factor influencing cost effectiveness results but the maximum duration of treatment likely in practice is unclear.

DTIC has not been included as a comparator in BRAF mutation-positive patients.

Summary of additional work undertaken by the ERG

The ERG conducted the following additional scenario analyses:

- A series of one-way analyses choosing different types of survival models for treatment efficacy. This includes:

    - using the Weibull, lognormal, log-logistic and generalised gamma distributions to model TTP for nivolumab and the Gompertz distribution for DTIC and ipilimumab
    - using the exponential, Gompertz, log-logistic, lognormal and Weibull distributions to model PFS for BRAF inhibitors (vemurafenib assumed to be equivalent to dabrafenib)
- Using the data extrapolation method to model long-term survival for nivolumab
- Including DTIC as a comparator in BRAF mutation-positive patients
- A scenario that combines the following assumptions:
    - using the Weibull distribution to model TTP for nivolumab patients
    - modelling PFS using the lognormal distribution for BRAF inhibitors
    - using the data extrapolation method to model long-term survival for nivolumab
    - between two years and no maximum treatment duration for nivolumab

Generally, the individual scenario analyses had a small impact on the base case model results, with changes to the method for estimating long-term overall survival for nivolumab (using data extrapolation) having the largest impact. This increased the ICER for nivolumab compared to ipilimumab in BRAF mutation-negative patients to £36,072 per QALY and in mutation-positive patients to £27,171 per QALY. The results of the combination scenarios had a much greater impact on the model results which showed nivolumab was dominated by ipilimumab in both the BRAF mutation-negative and BRAF mutation-positive patient groups.

The ERG repeated all the above analyses with the confidential PAS discounts for the comparator drugs in a separate confidential appendix for the NICE Appraisal Committee.

1 Introduction to the ERG Report

This report is a critique of the company's submission (CS) to NICE from Bristol Myers Squibb Pharmaceuticals Ltd on the clinical effectiveness and cost effectiveness of nivolumab for advanced (unresectable or metastatic) melanoma. It identifies the strengths and weaknesses of the CS. Expert clinical advice was sought by the ERG to inform this report.

Clarification on some aspects of the CS was requested from the manufacturer by the ERG via NICE on 17th September 2015. A response from the company via NICE was received by the ERG on 2nd October 2015 and this can be seen in the NICE committee papers for this appraisal.

2 BACKGROUND

2.1 Critique of the company's description of the underlying health proble

The CS generally provides a clear and accurate overview of the condition in sections 3.1 (CS p. 28) and 3.3 (CS p.33). However, the ERG notes that no reference is made to the genetic mutation BRAF (V600) that is prevalent in around 50% of people with melanoma. The presence of a BRAF mutation influences treatment choices.

Melanoma is described as an aggressive type of skin cancer which represents only 4% of all skin cancers, but accounts for 90% of skin-cancer related deaths. It mainly affects people of working age, with a mean age at diagnosis of 50 years. Incidence rates have been increasing over the past 50 years and the CS states that they are expected to continue rising by around 3.5% annually.

The CS estimates that 11,763 new cases were expected in England in 2013. Up to 10% of people diagnosed with melanoma present with advanced disease (unresectable or metastatic melanoma), and this is the patient group defined in the NICE final scope. The CS estimates there will be 1,304 new cases of advanced (unresectable or metastatic) melanoma in England during 2016. The company explains that the prognosis in advanced melanoma is generally poor and that life expectancy is commonly estimated at less than one year from diagnosis, but may have improved recently due to the availability of new treatments.

The CS lists a number of factors that can increase the risk of developing melanoma, and also lists prognostic factors (CS section 3.1 p 28).

2.2 Critique of company's overview of current service provision

The CS generally provides a clear and accurate overview of current pharmaceutical treatment options available to people with advanced melanoma in section 3.2 (CS p. 30). A list of relevant NICE guidance and other clinical guidelines is provided in Section 3.4 (CS p. 34). The company accurately describes current first-line treatments in advanced melanoma that have been recommended by NICE, with ipilimumab being the drug of choice for BRAF negative patients, and either ipilimumab or a BRAF inhibitor (either vemurafenib or dabrafenib) for those who have BRAF mutation. In this latter patient group, both drugs also represent second line treatment options for patients who did not receive them as their first-line therapy. For patients for whom ipilimumab or a BRAF inhibitor are not suitable, dacarbazine (DTIC) chemotherapy is the most common treatment in England. Last line systemic treatment is described as "palliative chemotherapy" regardless of BRAF mutation status. The ERG notes that this information is in line with current NICE guidlines[1], although the CS makes no mention of non-pharmacological options and service provisions described in the NICE Guideline NG14.

The CS does not explicitly describe which factors might make ipilimumab or a BRAF inhibitor unsuitable, or the proportion of patients this might apply to. After a clarification question from the ERG (clarification question A4) the company stated that eligibility for ipilimumab is determined according to the patient's overall fitness and the speed and extent of the disease. It is stated that patients should be fit enough to receive all four cycles of ipilimumab over a 12 week period. Expert clinical advice to the ERG also suggested that patients with immune toxicity (e.g. affecting people with rheumatoid arthritis) would be unlikely to be able to tolerate ipilimumab. The company did not state what factors might make treatment with a BRAF inhibitor unsuitable (other than BRAF mutation-negative status). However, expert clinical advice to the ERG suggested that there would be very few BRAF mutation-positive patients unable to take a BRAF inhibitor. The company also stated that of the BRAF mutation-negative population (who comprise 50% of the advanced melanoma population) up to 20% would not be suitable for ipilimumab and therefore may receive palliative chemotherapy, based on advice from UK clinicians participating in a BMS advisory board meeting. In summary, this appears to suggest there would be a minority of patients in whom ipilimumab or BRAF inhibitors would be unsuitable, and who, based on current management, would receive palliative chemotherapy or best supportive care (a comparator to nivolumab – see below).

The CS provides an overview of the limitations of current pharmacological treatment options in CS Table 7 (p. 35), stating that no long-term survival benefit has been demonstrated for BRAF inhibitor therapy or for chemotherapy (including DTIC chemotherapy). It is stated that the long-term survival benefit from ipilimumab treatment is observed in only 20% of patients. The CS suggests that the role of nivolumab in the clinical pathway will be to provide additional first and subsequent line treatment options that can be used regardless of BRAF status, and that are expected to provide longer-term survival benefits than currently available drugs (CS p. 32). Expert clinical advice to the ERG suggests that nivolumab could be potential a first-line treatment, in place of ipilimumab.

2.3 Critique of company's definition of decision problem

Population

The population is defined in the company's description of the decision problem as adults with advanced (unresectable or metastatic) melanoma. This is the population specified in the final scope issued by NICE and the ERG believes that this population is appropriate for the potential use of nivolumab in the NHS.

Intervention

The intervention described in the company's decision problem is nivolumab (brand name: Opdivo), and this is in line with the final scope issued by NICE. Nivolumab received marketing authorisation for advanced melanoma in June 2015. It is an immuno-oncology treatment that, according to the company, "stimulates the patient's own immune system to directly fight cancer cells" (CS p. 22).

As outlined in the CS (Table 2 p. 16, and chapter 2 p. 21 - 27), the summary of product characteristics (SmPC)[2] states that nivolumab as monotherapy is indicated for the treatment of advanced (unresectable or metastatic) melanoma in adults at a dosage of 3mg/kg every two weeks by intravenous infusion over 60 minutes. The treatment duration should be as long as there is clinical benefit or until treatment is no longer tolerated. The maximum duration is anticipated to be two years (CS Table 5, p. 25). Dose escalation or reduction is not recommended in the SmPC, and the company's European Public Assessment Report (EPAR) summary states that "dosing delay or discontinuation may be required based on individual safety and tolerability." Guidelines for treatment modifications and discontinuation are provided (CS Appendices Table 1 p. 4).

The CS states that the only contraindication is hypersensitivity to the active substance or to any listed excipients. However, the ERG notes that the CHMP has requested the

implementation of special warnings and precautions for the minimisation of immune-related adverse reactions that are associated with nivolumab treatment. The company describes these safety-related conditions of marketing authorisation on CS p. 25 and in Appendix 1, and these are specified in educational materials for professionals, patients and carers, including a "patient alert card" and a physician "adverse reaction management guide."

Overall, the intervention described in the decision problem reflects its use in the UK and is appropriate for the NHS. The impact on NHS service provision is described in CS section 2.4 (p. 26). The company points out that adequate infrastructure is already in place in the UK in the form of hospital oncology units, but adds that the nivolumab two-weekly dosing requirement represents a more frequent administration regimen than current therapies. The ERG notes that, in addition to the more frequent dosing of nivolumab as compared to current therapies, the continuous treatment of up to two years' duration may also impact on NHS service provision.

Comparators

The comparators of interest listed by the company are

- BRAF inhibitors (dabrafenib and vemurafenib – for people with BRAF V600 mutation-positive melanoma who have not previously received a BRAF inhibitor),
- Ipilimumab (for people who have not previously received ipilimumab),
- DTIC (for people who have received both a BRAF inhibitor and ipilimumab, or for whom either or both of these is/are unsuitable), and
- Best supportive care (for people who have received both a BRAF inhibitor and ipilimumab, or for whom either or both of these is/are unsuitable).

Referring to previous submissions to NICE, the company states that it considers DTIC to be a palliative chemotherapy, which forms part of best supportive care (CS Table 1, p. 14-15). Expert clinical advice to the ERG agrees with this and points out that the drug is rarely used in practice. However, the ERG notes that NICE Guideline NG14[1] recommends DTIC as a "systemic cancer treatment" for people with stage IV metastatic melanoma if immunotherapy or targeted therapy are not suitable. The NICE guideline adds in a footnote that "this use is common in UK clinical practice" but states that DTIC did not have a UK marketing authorisation for this indication at the time of guideline publication (July 2015).

The final scope specified DTIC as a comparator drug for patients who have previously received "both a BRAF inhibitor and ipilimumab, or for whom either or both of these is/are

unsuitable." The ERG notes that the economic analysis includes DTIC as a comparator for BRAF mutation-negative patients, but not for BRAF mutation-positive patients.

The CS does not refer to pembrolizumab (brand name: Keytruda) for the treatment of advanced melanoma in adults. NICE has recently recommended the use of pembrolizumab in advanced melanoma after disease progression with ipilimumab (NICE TA357),[3] and in patients not previously treated with ipilimumab (this recommendation is based on the final appraisal determination issued in October 2015. Final guidance is due in November 2015). The ERG notes that although pembrolizumab is a potential comparator to nivolumab it was not included in the final scope issued by NICE and the ERG therefore considers the company's choice of comparators to be appropriate.

Outcomes

The outcomes stated in the company's decision problem are all those specified to be of interest in the final scope:

- Overall survival,
- Progression-free survival,
- Response rate,
- Adverse effects of treatment,
- Health-related quality of life.

Economic analysis

The approach to the economic analysis proposed in the decision problem matches the final scope issued by NICE and is appropriate for the NHS. The company states that costs are considered from a National Health Service and Personal Social Services perspective, and that the availability of patient access schemes for the comparator technologies has been taken into account.

Other relevant factors

- Subgroups

The final scope does not specify any subgroups and the CS has not specified any subgroups in the decision problem. The ERG notes that the CS reports the results of various pre-defined subgroup analyses for the overall survival outcome from the CheckMate 066[4] trial in the main body of the CS (CS section 4.8 p. 89-91) and for CheckMate 067[5] in CS Appendix 7. The economic analysis presents results by BRAF mutation status. The ERG considers

this approach to be adequate although the usual caveats regarding subgroup analyses apply (e.g. small sample size, need for appropriate analysis, caution in interpretation).

- Equity or equality

No equity or equality issues were specified in the final scope, and the company did not identify any in their decision problem. The ERG is also not aware of any specific issues related to equity or equality in the use of nivolumab in patients with advanced melanoma, and expert clinical advice to the ERG confirmed that the more frequent dosing regimen required in nivolumab treatment compared to alternative treatments was unlikely to put patients at a disadvantage.

3 CLINICAL EFFECTIVENESS

3.1 Critique of the company's approach to systematic review

3.1.1 Description of the company's search strategy

The searches are generally fit for purpose, with the strategies well-constructed and with relevant search filters applied. An appropriate range of databases, including those recommended by NICE (Medline, Embase, Medline In-Process and Other Non-Indexed Citations, and The Cochrane Library), have been used, and tabulated, with only one minor transcription error. The search terms representing the indication were left broad (i.e. "melanoma") to maximise the number of references identified , rather than having been restricting to advanced, unresectable or metastatic disease.

The clinical-effectiveness searches, although deemed thorough with adequate documentation, contain three different searches:

(i) a search designed to identify RCTs of nivolumab and comparator therapies used in the first-line treatment of advanced melanoma, originally conducted in October 2014 and updated in May 2015;

(ii) a search to identify RCTs of nivolumab and comparator therapies in the subsequent-line setting, originally conducted in July 2014 and updated in May 2015.

(iii) a search aligned to the current decision problem, conducted in May 2015

This sequence of searches is assumed to be explained by the fact that originally there were three separate planned NICE single technology appraisals of nivolumab monotherapy for advanced melanoma, which were subsequently combined into the current appraisal.

The search strategies did not document the number of hits attained (returned) for each line of the strategy, which lessens immediate transparency and renders comparison of hits in replication of the searches more difficult.

The ERG replicated the Medline and Cochrane searches from the clinical-effectiveness search strategies as they were four months out of date (conducted on 7th/8th May 2015). No additional studies relevant to the systematic review of clinical-effectiveness inclusion criteria in the CS were identified from this search (CS Table 8, p. 38).

The ERG re-ran the searches for cost effectiveness, cost and resource identification and quality of life studies, since all three (dated 25th November 2015) were nine months out of date. No additional relevant studies were identified from this search, however, through *ad hoc* searching the ERG found identified a potentially relevant cost-effectiveness study reported in a 2015 conference abstract (see Section 4.1 of this report).[6] The School of Health and Related Research Health Utilities Database (ScHARRHUD) was additionally searched by the ERG, for utility papers on melanoma; however, the only reference found was already in the CS reference list.

Although the CS stated that annual proceedings of the conferences were hand searched in order to identify any relevant ongoing research (e.g. the American Society of Clinical Oncology), there were no specific details recorded of an ongoing trials search having been conducted on clinical trials databases. The ERG searched UKCRN, WHOICTRP, ISRCTN databases. One additional on-going trial was identified by the ERG (see Table 3 of this report)

Separate searches were undertaken for non-randomised studies of nivolumab (CS p. 121, CS Appendix 2.2.2). The CS states that these used similar methodologies and search strategies as those described for the systematic review of RCTs. The searches were conducted up to December 2014 for studies of nivolumab as first line treatment, and August 2014 for studies of subsequent line treatment. Given that these searches were for non-randomised studies the ERG has not updated them to the present time.

3.1.2 Statement of the inclusion/exclusion criteria used in the study selection.

The inclusion and exclusion criteria are clearly stated in CS Table 8 (p. 38). The inclusion criteria reflect the nature of the decision problem stated in the CS, the licensed indication,

and the current NHS position. Only randomised controlled trials (of any design type) were eligible for inclusion in the company's systematic review, and only those RCTs that investigated the clinical efficacy and/or safety of stated interventions were included (NB. The inclusion criteria included as ‘interventions’ all of the treatments listed in the decision problem, whether they were listed there as an intervention or a comparator, to permit an indirect comparison to be conducted – see Section 3.1.7 of this report). Systematic reviews and meta-analyses were included as a source of references. Inclusion criteria for outcomes were in line with the decision problem, but no exclusion criteria for outcomes were defined. The company explained that trials were not excluded on the basis of outcomes alone.

No limits were placed on inclusion relating to the quality of the RCTs, and setting was not used as an inclusion criterion.

A PRISMA flow diagram was provided showing the numbers of records included and excluded at each stage (CS Figure 6, p. 40). The diagram contains the numbers of records identified during the three database searches described above, as well as conference abstracts, three clinical study reports (CSR) and unspecified "other" eligible records.

The ERG notes that there is an unexplained discrepancy between the number of full-text articles assessed during the three database searches (n=240) and the number of unique full-text articles assessed for eligibility (n=204), possibly due to the removal of duplicates, for which no data were reported. All other sums are correct and a summary of the reasons for article exclusion was reported.

In total, 90 records of 44 studies were included and data sources for these were presented in CS table 9 (p. 41). A reference list of excluded reports (without reasons for exclusion) was provided in CS Appendix 5 (CS Appendices p. 64).

For non-randomised studies, a table of eligibility criteria is provided in CS Appendix 2.2 (CS Appendices p. 43). Only studies investigating nivolumab 3mg/kg monotherapy were eligible for inclusion, and other agents (e.g. the comparator drugs named in the decision problem) were excluded.

A PRISMA diagram for non-RCT evidence is also included (CS Appendices p. 45). All sums in the "first-line setting" searches are correct, but the ERG notes that there appears to be an error in the "subsequent-line setting" part of the diagram, where the number of records screened for eligibility (n=327) is smaller than the number of records subsequently excluded

(n=335). Seven records of two studies were included in the review of non-RCT evidence, but only the CheckMate 003 study[7] was subsequently discussed in the CS (CS p. 121). The other study was a phase I study[8] that did not provide additional data so this is not further discussed in the CS.

The company does not explicitly discuss bias, but states that the non-randomised CheckMate 003 study[7] was considered relevant to the decision problem because of its long-term survival data that support the company's position on nivolumab treatment duration and discontinuation. The ERG appreciates that long-term survival data from randomised studies of nivolumab are not yet available and considers the company's approach to providing supporting evidence from non-randomised studies to be reasonable. The ERG has not, however, reported the results of this study in detail in this report.

The ERG concludes that in general, inclusion and exclusion criteria for non-RCT studies are in line with the decision problem, the licensed indication and the NICE scope.

3.1.3 Identified studies

The CS identified and included three pivotal phase III RCTs of nivolumab monotherapy at the licensed dose in patients with advanced melanoma as specified in the NICE final scope. The trials (CheckMate 066[4], CheckMate 067[5] and CheckMate 037[9]) are reported in three journal articles and in six conference abstracts. All are international multi-centre studies, initiated in December 2012, (CheckMate 037[9]) January 2013 (CheckMate 066[4]) and June 2013 (CheckMate 067[5]). All are currently ongoing for extended follow-up. The company states that the CS used data from the CSR in addition to the published study results (CS p. 46).

The trials differ in their populations and comparators, as shown in CS Table 10 (p. 45):

- CheckMate 066[4] recruited treatment naïve, BRAF mutation-negative (wild-type) patients. The comparator in this trial was DTIC 1000mg/m2 administered every three weeks. The company explains that DTIC was the most common first-line therapy for BRAF mutation-negative patients prior to the approval of ipilimumab, and that this was the reason to include it in this trial as the comparator drug. In total, 418 patients were randomised (210 to nivolumab and 208 to DTIC as shown in CS Figure 7, p. 61).
- CheckMate 067[5] recruited treatment naïve patients with any BRAF mutation status. This was a three arm trial and the two comparator treatments were ipilimumab 3mg/kg administered every three weeks, and a combination of Nivolumab at a dose of 1mg/kg and ipilimumab 3mg/kg, administered every three weeks. The combination therapy

arm is outside of the NICE final scope and thus is not reported on in detail in the CS. The ipilimumab 3mg/kg arm of this trial allows a direct comparison between nivolumab and ipilimumab. A total of 945 patients were randomised, 316 to nivolumab and 315 to ipilimumab, as shown in CS Figure 8, p. 62. The remaining 314 patients were randomised to the combination therapy.

- CheckMate 037[9] recruited patients who progressed on or after prior anti-cytotoxic T-lymphocyte-associated protein 4 (CTLA-4) therapy (ipilimumab) and (if BRAF mutation-positive) BRAF inhibitor therapy. This was an open-label study with the comparator the investigator's choice of one of two chemotherapy options, either DTIC 1000mg/m2 or carboplatin area under the curve 6 + paclitaxel 175mg/m^2. Both comparators were administered every three weeks. In total 405 patients were randomised (272 to nivolumab and 133 to ICC (CS Figure 9, p. 63).

The ERG presents a summary of trial characteristics in Table 1.

Table 1 - Summary of characteristics of the included trials

	CheckMate 066 (n=418)	CheckMate 067 (n=631)[a]	CheckMate 037 (n=405)
Phase	Phase III	Phase III	Phase III
Blinding	Double blind	Double blind	Open label
Population	Previously untreated patients with advanced melanoma	Previously untreated patients with advanced melanoma	Previously treated patients with advanced melanoma
BRAF mutation status	Without BRAF mutation	With or without BRAF mutation	With or without BRAF mutation
PD-L1 status	PD-L1-positive, negative or intermediate classification	PD-L1-positive, negative or intermediate classification	PD-L1-positive, negative or intermediate classification
Comparator	DTIC	Ipilimumab	ICC
Primary outcome(s)	OS	OS, PFS	ORR, OS
Start date	January 2013	June 2013	December 2012
Status	Terminated[b]	Ongoing	Ongoing
Cut-off (database lock)	5 August 2014	17 February 2015	30 April 2014 (clinical database lock) 20 May 2014 (IRRC database lock)
Currently available primary/survival outcomes	1 year OS PFS	PFS	ORR PFS
Expected availability of further data	18 month OS: November 2015; 2 year OS: Q4 2016	OS and PFS: Q4 2016	OS and PFS: November 2015; OS extended follow/up: June 2016

DTIC = dacarbazine; ICC = investigator's choice chemotherapy (dacarbazine or carboplatin plus paclitaxel; IRRC = independent radiology review committee; ORR = objective response rate; OS = overall survival; PFS = progression-free survival; Q4 = quarter 4.

[a] Nivolumab monotherapy and ipilimumab monotherapy arms. The trial included a third arm of combined nivolumab and ipilimumab treatment, which not included in this ERG report.

[b] Recommendation by the data management committee to allow cross-over from DTIC to nivolumab treatment.

Reporting of study characteristics in the CS

The company submitted generally adequate summary details of the RCTs:

- Trial design, population (eligibility criteria), trial drugs (and permitted concomitant medications), outcomes (primary, secondary, and key exploratory outcomes), and pre-planned subgroups are described for all trials (CS Table 11, p. 49-55). Locations and settings are also included in this table.
- Patient numbers are shown in CS Figure 7 for CheckMate 066[4], Figure 8 for CheckMate 067[5] and Figure 9 for CheckMate 037[9] in the form of CONSORT diagrams (CS p. 61-63). Numbers of patients enrolled, randomised, and treated are provided, but the ERG notes that the numbers of patients screened for eligibility are not reported for any of the trials and no reasons are provided for loss of patients or exclusion of patients between enrolment and randomisation. All trials lost a small number of patients between randomisation and treatment and reasons for these withdrawals and exclusions are briefly discussed in the narrative summary of participant flow (CS p. 60).

  The numbers of patients who discontinued the trial medication during the course of the trial are reported in the CONSORT diagrams, and reasons are provided for discontinuations. All of these sums appear to be correct.

  The numbers of participants who continued to receive the study drug are reported for all included trials. The numbers of those who continued to participate in the study and are still being followed up for survival analysis are also reported. However, the ERG notes a possible error in the footnote attached to the CONSORT diagrams stating that "Continuing treatment means patients are continuing to receive study drug; continuing study means patients have discontinued study drug but are still being followed for survival analysis." In all CONSORT diagrams, the sum of patients continuing treatment and patients continuing study in the nivolumab group is larger than the number of patients treated with nivolumab. The ERG believes that the number of patients "continuing study" includes not only those who have discontinued the study drug, but also those who are continuing treatment.

  No numbers are reported for patients who crossed over between study drugs. The CS states (p. 46) that the option of crossing-over from DTIC to nivolumab for those who were not benefiting from treatment was permitted in the CheckMate 066[4] trial, after a study protocol amendment in June 2014 (approximately 18 months after trial initiation) was made in response to a recommendation by the data monitoring

committee. The company states that the data presented in the CS are based on a database lock dated 5th August 2014. At this time point no data on patients randomised to DTIC and subsequently treated with nivolumab post DTIC discontinuation were available (CS p. 46). The ERG supports the company's view that the results of the DTIC arm reported in the CS are unlikely to be confounded by un-blinding of treatment allocation or by subsequent nivolumab use.

- The methods of the statistical analyses of the nivolumab trials are summarised in CS Table 12 (p. 57-59). The table describes for each of the included trials the hypothesis objective, the statistical analysis, and the sample size and power calculations. Intention-to-treat (ITT) analyses were undertaken in all three trials for primary outcomes, and censoring methods were used to take account of missing data. The company's selection of outcomes is described in section 3.1.5 of this report.

The CS did not identify any specific patient groups in the decision problem for whom subgroup analyses were required. However, as stated earlier in this report, the comparative summary of RCT methodology (CS Table 11, p. 49-55) specified a range of pre-planned subgroup analyses for each trial, to assess the impact of participant characteristics (including demographic data and a range of disease-related baseline characteristics), and the geographic regions of the trials.

CSRs were supplied to the ERG by the company for information, though the ERG has not performed an analysis of these in the preparation of this report. All of the included studies of nivolumab were sponsored by the company.

Characteristics of study participants

Baseline characteristics of participants in the included RCTs are presented in CS Table 13 (p. 65-68). The company states that baseline characteristics of CheckMate 066[4] and CheckMate 067[5] are "well balanced with no key differences between treatment groups." Overall, the ERG agrees with the company's assessment, but notes that participants from both trials appear to be somewhat older in the comparator arms. The ERG also notes that patients randomised to DTIC in CheckMate 066[4] appear to have poorer Eastern Cooperative Oncology Group (ECOG) performance status as compared to those in the nivolumab group. However, expert clinical advice to the ERG suggested that the observed differences in ECOG performance status are unlikely to be clinically significant.

For the CheckMate 037 trial[9] the CS describes baseline characteristics as "generally well balanced." The company points out that higher proportions of patients with a history of brain metastases or with higher LDH were observed in the nivolumab group, suggesting that patients randomised to nivolumab had a poorer prognosis than those in the comparator group (CS p.64). The ERG also observed that the ECOG performance scores appeared to be somewhat lower in the nivolumab group. Overall, the ERG agrees with the company's assessment, but again, the significance of these observations remains unclear.

The ERG also agrees with the CS that differences between trials in patients' baseline characteristics are attributable to the individual trial eligibility criteria. The company points out that participants in CheckMate 037[9] were previously treated, and therefore had a longer time from diagnosis than those in CheckMate 066[4] and CheckMate 067.[5] They were also on average younger than those in CheckMate 066[4] and CheckMate 067,[5] and the company believes that this may reflect younger patients' ability to withstand multiple lines of therapy. The ERG notes that a higher proportion of CheckMate 037[9] participants (>75%) appeared to have metastasis stage M1c (the most severe M category) as compared to 61% in CheckMate 066[4] and 58% in CheckMate 067.[5] Expert clinical advice to the ERG points out that the CheckMate 037 participants are noteworthy as they have been able to receive several lines of treatment, but have other poor prognostic features such as higher M1c compared with the other trial populations.

Overall, the ERG agrees with the company that there are no noteworthy differences in patient characteristics between CheckMate 066[4] and CheckMate 067[5],and that differences in patient characteristics between these trials and CheckMate 037[9] are reflective of the fact that failure of previous treatments was an eligibility criterion for this trial.

In order to assess the applicability of the CheckMate trials to the UK patient population The ERG asked the company to confirm the number of UK participants in each trial and provide their baseline characteristics (clarification questions A1 and A2). In CheckMate 037[9] there were five UK trial centres and 43 UK patients were randomised to treatment. Seven UK trial centres participated in the CheckMate 067[5] study, with 93 UK patients randomised to treatment (27 to nivolumab, 36 to ipilimumab, and 30 to nivolumab in combination with ipilimumab). No UK patients were enrolled in CheckMate 066.[4] The baseline characteristics of UK participants were presented in appendices 1 and 2 of the company's response to the clarification questions (1st October 2015). There were some differences between UK patients and the total CheckMate 067 and CheckMate 037 trial populations, although most were small. Of note are differences in PD-L1 status and BRAF mutation status, presented in

Table 2. The proportion of participants with positive PD-L1 status in CheckMate 067 appeared to be higher in the UK group compared to the total trial population, and in CheckMate 037 BRAF mutation-positive status was found to be more prevalent in UK patients. The ERG is uncertain whether these differences are significant, given the small size of the UK patient group.

Table 2 – Differences in PD-L1 and BRAF status between UK participants and trial populations

	CheckMate 067		CheckMate 037		
	UK participants (n=93)	Total trial population (n=945)[a]	UK participants (N=43)	Total trial population (n=405)	
				Nivolumab	ICC
PD-L1 positive, %	49.5	23.6	46.5	49	50
BRAF mutation-positive %	36.6	31.5	34.9	22	22

[a] The company provided UK participant data across all three arms of the CheckMate 067 trial.

Ongoing trials

The CS identified five ongoing studies. Three of these (CheckMate 066, CheckMate 067, CheckMate 037) are the trials included in the CS[4;5;9]. These are currently ongoing or in extended follow-up in order to generate evidence on long-term outcomes, including overall survival, progression-free survival, and HRQoL (CS Table 53, p. 150). Two further studies mentioned in the CS are CheckMate 069, a phase II RCT of nivolumab in combination with ipilimumab compared to ipilimumab alone, and CheckMate 064, a phase II study that investigates the sequential administration of nivolumab and ipilimumab. Both trials are outside the NICE scope and decision problem defined for this CS.

The ERG notes that the company only listed ongoing trials that are expected to report data within the next 12 months. A search for ongoing trials undertaken by the ERG identified just one additional relevant study – a single-arm study of nivolumab in patients progressing after previous anti-CTLA-4 treatment (Table 3).

Table 3 - Ongoing trials

Trial identifier, sponsor	Design, Country	Intervention, comparator, patient group	Expected end date
NCT02156804 Bristol-Myers Squibb	Single-Arm, Open-Label, Multicentre Clinical Trial. International (168 sites, incl. 15 UK sites)	Nivolumab 3 mg/kg every two weeks. No comparator. Subjects with histologically confirmed stage III (unresectable) or stage IV melanoma and progression post prior treatment containing an anti-Cytotoxic T Lymphocyte Antigen (CTLA-4) monoclonal antibody (N=800)	October 2017 (Final data collection date for primary outcome measure).

In summary, all three the RCTs included in the systematic review of clinical effectiveness meet the inclusion criteria, and the ERG believes that it is likely that the CS has identified all relevant RCTs. The CS provides generally adequate details of the characteristics of the RCTs.

3.1.4 Description and critique of the approach to validity assessment

The company critically appraised the included nivolumab trials using the NICE-recommended criteria and presents a summary of findings on CS p. 69 and in CS Table 14 (p. 70). The complete quality assessments of each of the RCTs are included in CS Appendix 3 (CS Appendices p. 55, Tables 10-12). The ERG agrees with the company assessment for most criteria (Table 4).

The ERG assessment differs for question 1 (randomisation) because the sequence generation process is not described. The ERG notes that stratified allocation methods were applied in the randomisation procedures for all three trials. In CheckMate 066[4], randomisation was stratified by PD-L1 status and metastasis stage via permuted blocks within each stratum. In CheckMate 067[5] randomisation was also performed by permuted bocks within strata, and stratification was defined by PD-L1 status, BRAF mutation status and metastasis stage (as per American Joint Committee on Cancer definition). In CheckMate 037[9] randomisation was stratified by PD-L1 status, BRAF mutation status and prior anti-CTLA-4 best response.

For question 3 (balance in prognostic factors) the ERG notes small imbalances between groups in CheckMate 066[4] (relating to age and ECOG PS scores 0 and 1) and CheckMate 067[5] (relating to age) as described above (section 3.1.3 of this report and CS Table 13 p. 65-68). The potential impact of these small imbalances on trial outcomes is not clear; however, age and ECOG PS score are two of the baseline characteristics with known prognostic effects on outcomes (presented in CS Table 26 p. 101).

Table 4 - Company and ERG assessment of trial quality

		CheckMate 066[4]	**CheckMate 067[5]**	**CheckMate 037[9]**
1. Was randomisation carried out appropriately?	CS:	Yes	Yes	Yes
	ERG:	Not clear	Not clear	Not clear
Comment: Randomisation was stratified in all of the trials. In CheckMate 066 and 067 randomisation was performed by permuted blocks within each stratum, as described in the study				

		CheckMate 066[4]	CheckMate 067[5]	CheckMate 037[9]
protocols (supplementary material published online). The ERG notes that the sequence generation process (e.g. use of a random number table or random number generator) is not described.				
2. Was concealment of treatment allocation adequate?	CS:	Yes	Yes	Yes
	ERG:	Yes	Yes	Yes
Comment: Randomisation was performed by interactive voice response system.				
3. Were groups similar at outset in terms of prognostic factors?	CS:	Yes	Yes	No
	ERG:	Not clear	Not clear	No
Comment: The ERG notes small imbalances between groups in CheckMate 066 (relating to age and ECOG PS scores 0 and 1) and CheckMate 067 (relating to age).				
4. Were care providers, participants and outcome assessors blind to treatment allocation?	CS:	Yes	Yes	Outcome assessors only
	ERG:	Yes	Yes	Outcome assessors only
Comment: Use of matched placebos in CheckMate 066 and 067. CheckMate 037 is an open-label study, where patients and care providers were not blind to treatment allocation. Primary efficacy assessment of ORR was conducted by an independent radiological review committee, and committee members were blind to treatment allocation.				
5. Were there any unexpected imbalances in drop-outs between groups?	CS:	No	No	Yes
	ERG:	No	No	Yes
Comment: Although higher proportions of patients discontinued the study treatment in the comparator groups of CheckMate 066 and 067 this was due to greater proportions discontinuing due to disease progression. Discontinuations for other reasons were similar between groups. In CheckMate 037 a number of patients randomised to ICC withdrew consent, resulting in an imbalance in numbers of patients withdrawing between groups prior to treatment initiation. The company explains that withdrawals included patients who went on to receive other PD-1 therapies outside of the trial, and this would have had an impact on the outcome of OS of the ITT population.				
6. Is there any evidence that authors measured more outcomes than reported?	CS:	No	No	No
	ERG:	No	No	No
Comment: In CheckMate 067 OS data were not yet available when the CS was produced.				
7. Did the analysis include an ITT analysis? If so, was this appropriate and were appropriate methods used to account for missing data?	CS:	Yes	Yes	Yes
	ERG:	Yes	Yes	Yes
Comment: ITT analyses were performed for the primary outcomes in all of the trials and outcomes were censored on the last date the subject was known to be alive (for OS) or on the date of the last tumour assessment (for PFS). The ERG considers this approach to be appropriate. In CheckMate 037 the approach to censoring only appeared to be reported for the time to response outcome rather than other time to event outcomes.				

3.1.5 Description and critique of company's outcome selection

The outcomes selected in the decision problem match the NICE scope and are appropriate for the assessment of cancer drugs.

Overall survival (OS) and progression free survival (PFS) were defined consistently between the three key RCTs included in the CS. Response is defined as 'objective response rate' (ORR) in all three trials, consisting of the 'best overall response' (BOR) of complete or partial response (CR or PR) divided by the number of randomised patients. (IRRC as well as investigator assessed in CheckMate 037, where it was the primary outcome measure).

Response was measured by Response Evaluation Criteria in Solid Tumors (RECIST) criteria[10] (version 1.1) in all three trials. Time to treatment response (TTR) was reported in all three trials and defined consistently (IRRC as well as investigator assessed in CheckMate 037). Duration of response (DOR) is also reported for all three trials, and defined consistently between them (IRRC as well as investigator assessed in CheckMate 037).

HRQoL was measured using the European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire-Core 30 (EORTC-QLQ-C30) scale, as a secondary outcome in the three trials. In addition the EQ-5D, and the Work Productivity and Activity Impairment Questionnaire: General Health (WPAI:GH) were used to measure HRQoL as exploratory outcomes in the trials (WPAI:GH was not measured in CheckMate 037).

Adverse events were measured in all three trials, including deaths and laboratory abnormalities. Severity was measured using the National Cancer Institute Common Terminology Criteria for Adverse Events.

In terms of instrument validation, the EQ-5D has been validated and used in a number of economic evaluations. However, the CS does not mention the EQ-5D value set used in the trials and it is therefore unclear how applicable the preference weighting is to the UK general population. The CS does not state whether the EORTC-QLQ-C30 scale is validated (CS Appendix 6 provides limited further information on this instrument). The EORTC website reports that it is has been translated and validated into 81 languages and is used in more than 3,000 studies worldwide, though this does not necessarily imply scientific validation (http://groups.eortc.be/qol/eortc-qlq-c30). The website also mentions that a melanoma module is under development – called QLQ-MEL38.[11] (not mentioned in the CS).

There are no additional outcomes reported in the clinical trial publications that are not included in the CS. However, the ERG notes that TTP, PrePS and PPS outcomes are not reported in the trial journal publications, but are presented in the CS specifically to inform the transitions in the economic model (see Section 3.1.7 and Section 4.2.4 of this report).

3.1.6 Description and critique of the company's approach to trial statistics

The CS reports all relevant outcomes for the three included primary RCTs, apart from overall survival and HRQoL which are only reported for CheckMate 066. The CS states this is because OS data for the other two RCTs are currently immature due to insufficient follow-up,

whilst HRQoL data for CheckMate 067 and 037 are likely to become available in the next 12 months.

The outcomes are classed as primary, secondary and exploratory (CS Table 11, p.54). Primary outcomes are those for which the trials are powered statistically (the meaning of secondary and exploratory outcomes is not defined in the CS). Primary outcomes are OS in all three RCTs, plus the co-primary outcomes of PFS in CheckMate 067 and ORR in CheckMate 037. Secondary outcomes vary across the three RCTs and include: PFS and ORR (CheckMate 066, 067), TTR, DOR and PFS (CheckMate 037 only) and HRQoL (measured in all three RCTs, but reported in the CS for CheckMate 066 only). Secondary outcomes also include OS for specified subgroups, which in all RCTs include presence or absence of PD-L1 expression. The exploratory outcomes are adverse events in all three RCTs and, in CheckMate 067, also the time to TTR and DOR.

Two RCTs tested hypotheses of superiority of nivolumab against dacarbazine (CheckMate 066) or ipilimumab monotherapy (CheckMate 067) whilst CheckMate 037 also appears to have tested a superiority hypothesis although this was phrased as “nivolumab will provide meaningful activity” compared to ICC (CS Table 12, p. 57-59).

All three RCTs randomised a larger number of participants than their intended sample sizes (CS Table 12, p. 57-59). However, due to the interim nature of the reported analyses, which is acknowledged in the CS, the number of events (death or progression) which had occurred by the time of analysis of primary outcomes were fewer than the number required to achieve the statistical power specified in the CS for detecting pre-specified HRs for overall survival or for progression-free survival (CS Table 12, p. 57-59). For example, CheckMate 066 required ≥312 deaths to detect a pre-specified HR for overall survival of 0.69 at 90% power (2-sided α=0.05) but at the time of analysis only 146 deaths in total across both trial arms had occurred (CS Table 15, p.72). Sources of some assumptions in the power calculations are not explained in the CS (CheckMate 066 assumed median OS of 10 months for DTIC and 14.49 months for nivolumab; CheckMate 067 assumed median OS of 14 months for ipilimumab and 19.4 months for the comparator arms). However, on balance, given that the analyses in these RCTs are testing superiority rather than equivalence, the ERG believes that the under-powering of these interim primary outcome analyses would not influence interpretation of the reported analyses. As stated in the CS, further follow-up data for these analyses will be reported during 2015-2016.

The statistical analyses for determining time-to-event measures (OS, PFS, ORR, DOR, and TTR) in each of the RCTs are based on standard Kaplan-Meier survival analysis methods (CS Table 12, p. 57-59). Outcomes are reported as event rates and as median values with 95% CIs (several approaches for calculating the 95% CIs are reported in the CS based on published methods).

In CheckMate 066 and 067, comparisons of survival across treatment arms are based on Cox proportional hazards models to give hazard ratios for death or disease with 95% CI, median OS with 95% CI, and median PFS with 95% CI (CS p. 73-76). Comparisons of ORR across trial arms were estimated using Cochran-Mantel-Haenszel tests to calculate odds ratios with 95% CI in CheckMate 066 (CS Table 17, p. 77) and CheckMate 067 (Table 18, p. 81). In CheckMate 037 ORR are compared across trial arms as differences in rates with two-sided 95% CI (Newcombe approach) (CS Table 19, p. 84).The CS states that analyses were stratified by prognostic variables: metastasis stage and PD-L1 status in CheckMate 066 for analyses of OS, PFS and ORR; and metastasis stage, PD-L1 status and BRAF status in CheckMate 067 for analyses of PFS and ORR (CS Table 12, p. 58). Subgroup analyses are also reported for these prognostic variables as well as for a range of other prognostic and demographic variables for OS in CheckMate 066 (CS Figure 22), for PFS in CheckMate 067 (CS Appendix 7) and for ORR in CheckMate 037 (CS Appendix 7).

Secondary outcomes derived from the Kaplan-Meier analyses and reported in all 3 RCTs are the median TTR plus range (CS Tables 17-19, p. 77-84) and the median duration of response plus range (CS Tables 17-19, p. 77-84) (also reported with 95% CI for CheckMate 066: CS Figure 14, p. 78).

All three RCTs employed an ITT analysis approach (Table 4). The method of data censoring was reported for progression-free survival in CheckMate 066 and 067, for overall survival in CheckMate 066, and for time to response in CheckMate 037 (CS Table 12, p. 57-59) but not for the other time-to-event outcomes reported in these trials.

Analysis of HRQoL in CheckMate 066 was based on the EORTC QLQ-C30 instrument and EQ-5D using a Cox proportional hazards regression model to determine time to first deterioration and first improvement (as defined by the minimal important difference for each scale applied at individual patient-level). Results are presented as hazard ratios for nivolumab versus DTIC with 95% CI (CS Table 20, p. 88-89).

Adverse events in all 3 RCTs are reported as numbers (%) of events, numbers (%) of discontinuations, and the median times (without variance measures) to onset and to resolution of events. For CheckMate 067 the median number of adverse events with 95% CI was estimated from Kaplan-Meier survival analysis for time on treatment (CS Figure 46, p. 139) whilst for CheckMate 037 a Kaplan-Meier survival curve for time on treatment is presented but without accompanying statistics (CS Figure 47, p. 142).

Overall, the ERG believes that the statistical analysis approaches employed in the 3 RCTs are generally appropriate. Where survival analyses were employed, the resulting curves are clearly reported in Figures together, in most cases, with the derived statistical parameters. However, the method of data censoring was not reported for the primary outcomes in CheckMate 037 and the ERG noted that this trial also had unexpected unbalanced attrition (see Table 4).

Company's approach to trial statistics in non-randomised studies

The single non-randomised study included in the CS, CheckMate 003, was a phase I study of nivolumab safety in treating solid tumours, including melanoma (CS Table 42, p. 123-124). Participants with melanoma were assigned across five dose cohorts (0.1, 0.3, 1.0, 3.0 and 10.0 mg/kg nivolumab every 2 weeks). Initially, small numbers of patients were allocated to the dose cohorts but maximum tolerated dose was not reached and "expansion cohorts" of further patients were allocated to the 3.0 and 10.0 mg/kg cohorts, as well as further patients randomly allocated to the 0.1, 0.3 and 1.0 mg/kg groups. Overall sample size at analysis was N=107 melanoma patients in total across all dose cohorts, of which n=17 were in the licensed dose cohort (3.0 mg/kg). Although described as a dose escalation study, dose changes were not permitted for individual patients unless allocated to the 0.1 or 0.3 mg/kg expansion cohorts who could escalate to 1.0 mg/kg if disease progressed within the first two treatment cycles. The flow of participants in this study from enrolment to analysis is not explicitly reported and as such is difficult to follow – the CS refers to "participant flow" when citing Table 43 (CS p. 126) but this merely presents a cross-sectional overview of patient status at analysis.

Safety and tolerability were specified as the primary outcomes in CheckMate 003. Secondary outcomes are listed in the CS as immunogenicity, pharmacokinetics, "preliminary efficacy", and characterisation of the dose-response relationship in melanoma (and non-small-cell lung cancer). The specified secondary efficacy outcomes included objective response rate, progression-free survival, duration of response and time to response, while overall survival was specified as being an exploratory outcome. The CS does not define

what are meant by primary, secondary or exploratory outcomes, and no mention is made in the CS of the statistical power of the study to detect effects on any outcomes.

The ORR and stable disease rates in CheckMate 003 were estimated together with 95% CI by using the Clopper–Pearson method. Time-to-event end points, including progression-free survival, overall survival, and duration of response were estimated by using Kaplan-Meier survival analysis methods, with 95% CI based on Greenwood's formula. Survival data were collected retrospectively. The CS states only that efficacy analysis was based on all treated patients with standard censoring methods to account for missing data, without providing details (CS, p. 125). Analyses are reported for different database lock times for each outcome, although it is unclear how the data availability at each analysis time relate to the cross-sectional overview of patient status at analysis as reported in CS Table 43 (p. 126).

Outcomes are reported in CheckMate 003 as ORR without variance measures or survival curves for the overall population and the licensed dose cohort (CS p. 128-130); median PFS with 95% CI for all five dose cohorts combined (CS Fig 43, p. 132); and median OS with 95% CI (CS Fig. 44, p. 133) for all five dose cohorts combined and for the licensed (3mg/kg) dose cohort alone.

Overall, the ERG believes that the results of CheckMate 003 should be interpreted with caution, due to the small sample size in the relevant dose cohort (n=17 only), uncertainty about relevance of the analyses on the overall study population (since these included non-licensed nivolumab doses); and lack of clarity regarding participant flow in relation to analysis timing and data censoring.

3.1.7 Description and critique of the company's approach to the evidence synthesis

Narrative synthesis

A narrative review of the nivolumab RCTs is provided. Each outcome measure is taken in turn (e.g. survival analysis, response analysis) with tabulated data and Kaplan-Meier survival curves and other figures provided for each of the three key trials respectively. A narrative description accompanies the tables and figures. The ERG has cross-checked the outcome data in the CS, where available, with that provided in the trial journal publications, and these are consistent, with only one identified exception – target lesion reduction in the CheckMate 066 trial. The CS (p. 80) reports that of the 103 patients treated beyond Response Evaluation Criteria in Solid Tumors (RECIST)[10] defined progression (54 nivolumab; 49 DTIC), 12 (22.2%) treated with nivolumab and 2 (4.1%) treated with DTIC developed or

maintained a target lesion reduction of >30% compared to baseline. In the trial journal paper[4] the corresponding figures are 17 (31%) and 8 (16%). Note that the journal paper states ‘a reduction of 30% or more’ in the target lesion whereas the CS just says ‘>30%’, which may explain the discrepancy.

A meta-analysis was not conducted, as the CS states that the clinical trials are too clinically diverse to be combined (CS section 4.9, p. 92). The key reasons include differences in control arms (DTIC, ICC and ipilimumab) and differences in patient populations enrolled (e.g. previous treatment experience; BRAF mutation status - though the ERG notes that much of the CS analysis assumes no independent effect of these patient variables anyway). The ERG agrees with the rationale for not meta-analysing, primarily due to differences between the trials in the comparator drug.

Indirect comparison overview

It is stated (CS Section 4.10, p.97) that a mixed treatment comparison of all the treatments within the scope of the appraisal was not possible for a number of reasons, including non-proportional hazards between the different drugs due to their differing mechanisms of action; cross-over of patients in some but not all of the trials; and heterogeneity in the trial designs (e.g. in terms of previous treatment experience, and BRAF mutation status). The ERG agrees that a mixed treatment comparison would be difficult to construct and interpret due to these reasons. However, the ERG notes that the company have made an apparent contradiction in their subsequent indirect comparison by assuming that previous treatment status and BRAF mutation status do not independently influence treatment effects (CS p. 100; see below for more detail).

The company reports an indirect comparison of nivolumab with its comparators (CS section 4.10). A ‘broad evidence’ network diagram is presented (CS Figure 23) showing the treatment comparisons possible from the trials of DTIC, dabrafenib, vemurafenib, ipilimumab and nivolumab that met the inclusion criteria for the company’s systematic review of clinical effectiveness (n=44) (CS section 4.1). It is stated that only trials that reported OS were eligible for inclusion in the indirect comparison as this is considered to be the most important outcome in patients with advanced melanoma. The ERG agrees with this assertion but also notes that PFS is also a clinically relevant outcome measure.

It is not stated in the CS how many trials were ineligible on this criterion but it does apply to the pivotal CheckMate 067 RCT for which the OS data are stated to be currently unavailable as the required minimum follow-up has not yet been reached. The ERG considers this to be

a significant omission as it would obviate the need for an indirect comparison of nivolumab and ipilimumab since they were compared head-to-head in this trial.

A network diagram is presented showing the comparisons between the trials eligible for the indirect comparison (CS Figure 24). There were five such trials included:

- CheckMate 066 (nivolumab vs DTIC)[4],
- BRIM-3 (vemurafenib vs DTIC)[12],
- BREAK-3 (dabrafenib vs DTIC)[13],
- CA184-024 (ipilimumab 10mg/kg + DTIC vs DTIC)[14], and
- MDX010-20 (ipilimumab vs ipilimumab 3mg/kg + gp-100 vs gp-100)[15].

As described below, three of these trials are subsequently used to inform the analysis (CheckMate 066[4], BRIM-3[12] and MDX010-20[15]).

Two indirect comparison networks were analysed, differing according to the type of comparators used:
(i) comparison with ipilimumab and palliative chemotherapy;
(ii) comparison with BRAF inhibitors.
Each of these is described and appraised in turn below, in terms of the identification of the clinical trial evidence used to conduct the indirect comparisons and the statistical procedures used (e.g. covariate adjustment to account for differences between trial arms). Section 4.2.4 of this report describes and critiques the statistical procedures used to fit and extrapolate parametric survival curves from the trials to inform the comparisons made within the economic model. As described in the following sub-sections, the indirect comparison used an approach whereby selected trial arms were compared using a covariate-adjusted survival model approach. This nomenclature is used in the CS and in this report to distinguish it from an adjusted indirect comparison that the company also reported, for purposes of comparison (described below).

(i) Indirect comparison of nivolumab to ipilimumab and palliative chemotherapy
This comparison informed the cost-effectiveness analysis for BRAF mutation-negative patients and comprises comparisons of treatments from trials using a common comparator. CS Table 25 describes the comparisons made. For nivolumab compared to ipilimumab, patient-level data from the CheckMate 066 trial[4] (nivolumab arm) were compared to patient-level data from the MDX010-20 trial[15] (ipilimumab arm) linked together by DTIC (CheckMate 066) and by gp100 melanoma peptide vaccine (MDX010-20) (NB. gp100 is assumed to be equivalent to DTIC in efficacy and therefore is used as a proxy for DTIC for purposes of

comparison – see below). Figure 1 illustrates the evidence network used (replicated from CS Figure 26). The CA184-024 trial[14], which used a higher dose of ipilimumab (10mg/kg) was only used in a scenario analysis as the CS states that 3mg/kg of ipilimumab and 10 mg/kg cannot be assumed to be equivalent (as noted in NICE TA319[16]). Nivolumab was compared directly to DTIC using patient-level head-to-head data from the CheckMate 066 trial.[4]

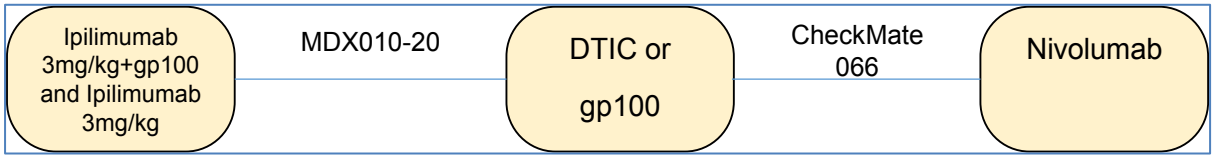

Figure 1 - Network diagram for the comparison of nivolumab with ipilimumab

The company makes the following assumptions for the indirect comparison:

1. *The line of treatment does not independently predict treatment effectiveness*. CheckMate 066 only included previously untreated patients, whilst MDX010-20 included patients who had been previously treated. The CS cites studies of ipilimumab and nivolumab that support this assumption and reports that this assumption was accepted in the NICE TA319 of ipilimumab.[16] Clinical advice to the ERG agreed with this assumption.
2. *There is no difference in treatment effect by BRAF mutation status.* This assumption was necessary because CheckMate 066 included only BRAF mutation-negative patients, MDX010-20 did not report the BRAF mutation status of patients. To support this assumption the CS cites a published retrospective pooled analysis of four on-going nivolumab studies by Larkin and colleagues (2015)[17], sponsored by the company. The ERG has done a brief assessment of this study and notes that three were phase I studies of nivolumab (respective sample sizes <100 patients), and the fourth was the phase III CheckMate 037 RCT[9], described earlier in this report. Of the 440 patients analysed, 334 were BRAF mutation-negative and 106 were BRAF mutation-positive, and 83% of the patients received nivolumab at the licensed 3mg/kg dose. The outcome measure used in the analysis by Larkin and colleagues[17] was treatment response though the CS uses survival in the indirect comparison, and it is not clear whether the assumption made on the basis of response analysis is necessarily applicable to survival. Limited details are provided of the included studies or the analysis methods to pool the studies and due to its retrospective nature the ERG urges caution in the interpretation of its results, and therefore its use to support the assumption.
3. *Gp100 is equivalent to DTIC in terms of OS and PFS outcomes*. The CS provides a rationale for this assumption that citing published meta-analyses[18;19] (both of which appear to be sponsored by BMS) of gp100 and existing treatments, including palliative

chemotherapy, showing them to be similar for OS. The CS also states that this assumption had been discussed and accepted in NICE appraisals of ipilimumab (NICE TA268[20] and TA319[16]). Expert clinical advice to the ERG agreed that these drugs can be considered generally equivalent and also that DTIC can be considered as palliative chemotherapy. However, the ERG notes that the CS does not report any evidence for the equivalence of gp100 and DTIC (or other palliative chemotherapy) for alternative cancer outcomes, including those used to inform transition probabilities in the economic model: time to progression (TTP), post-progression survival (PPS) and pre-progression survival (PrePS) (see below for discussion of these outcomes). The ERG notes that the Kaplan-Meier curves for TTP (measured from day 100) for gp100 and for DTIC do not appear to be similar (though this comparison is unadjusted and is based on small numbers of patients remaining in the trial arms). It is therefore unclear whether the equivalence of DTIC and gp100 can also be demonstrated for these outcomes.

4. *Ipilimumab 3mg/kg + gp100 and ipilimumab 3mg/kg are equivalent based on the MDX010-20 results.* The CS notes that this was an accepted assumption in the NICE TA268 of ipilimumab in previously treated patients, and this therefore allowed the pooling of these two trial arms to provide a larger dataset for analysis than if only the ipilimumab 3mg/kg monotherapy arm of MDX010-20 had been used.

Alternative outcomes

The CS states that the OS data for CheckMate 066 are relatively immature (i.e. they do not reach median survival) and long-term survival extrapolations of OS will therefore be subject to uncertainty (CS p. 101). The CS used the following alternative outcomes to inform long-term extrapolations. These were:

- Time to progression (TTP) – similar definition to PFS, however patients classified as progressors in PFS due to death are censored at death.
- Pre-progression survival (PrePS) – the same definition as OS except patients that progress are censored at time of progression.
- Post-progression survival (PPS) – only included patients that have progressed and follows time to death, or censoring, from the point of progression.

TTP and PrePS were used to inform long-term extrapolations of PFS. TTP, PrePS and PPS were used to inform long-term extrapolations of OS.

The ERG notes that these outcomes were not pre-specified as primary or secondary outcome measures for the CheckMate RCTs and data for them were not provided in the

main clinical effectiveness results section of the CS (CS Section 4.7, p. 71) or in the trial journal publications. (NB. The data are given in CS Section 4.10 'Indirect and mixed treatment comparisons'). They therefore appear to have been used retrospectively for the purposes of informing the economic model for this appraisal.

The TTP survival data are also split into two time periods (pre-and post-100 days) which use different modelling methods. This was done to allow a more clinically and statistically plausible shape and continuous flow to the occurrence of progression from day 100 onwards (CS p. 104; for a more detailed description and critique of this please refer to Section 4.2.4 of this report).

The CS argues that, due to the immaturity of OS survival data for nivolumab, use of the alternative outcomes allowed a more robust estimation of long-term survival extrapolations in (CS p. 101). This ERG acknowledges that this approach does avoid using immature OS data, but by using three endpoints rather than two, and splitting one of these (TTP) into two time periods, the sample sizes become smaller and the attendant survival curves are based on smaller samples and will have fewer observed events. To this extent they will also be less robust. For example, the Kaplan-Meier curves for PrePS (CS Figure 34) shows a population at risk in the nivolumab arm of 210 at outset, but of only seven at approximately 12 months, with apparently no observed events between six months and 12 months (as the curve for nivolumab is flat between these time points). Furthermore, median survival for PrePS is not reached for nivolumab or DTIC suggesting data immaturity for this outcome (CS Figure 34).

Covariate adjustments for the parametric survival model indirect comparison

To account for potential differences in patient characteristics between the CheckMate 066[4] and MDX010-20[15] trials the CS identified factors shown by a meta-analysis of trials by Korn and colleagues[21] to affect prognosis (in terms of OS and PFS) in patients with advanced melanoma treated with palliative chemotherapy. The CS applied the prognostic factors from the Korn and colleagues[21] meta-analysis to the TTP, PrePS and PPS outcomes which inform the economic model (see below). These factors are reported to be consistent with prognostic factors used in NICE TA319 of ipilimumab in previously untreated advanced melanoma.[16]

CS table 26 illustrates the comparability of the CheckMate 066[4] and MDX010-20[15] trials in terms of seven prognostic factors (six of which were baseline patient characteristics). The list of factors was reported to have been validated with UK clinicians during an advisory board

meeting in March 2015. Clinical advice to the ERG indicated that there were no key prognostic factors absent from those chosen.

There were differences between the trials in certain factors:

- Eastern Cooperative Oncology Group (ECOG) performance status zero (higher CheckMate 066)
- M stage disease (extent of metastatic melanoma) = M1c (higher in the MDX010-20 trial indicating more visceral disease)
- History of brain metastases (higher in the MDX010-20 trial)
- Age (higher in the MDX010-20 trial)
- Subsequent ipilimumab use (occurred only in CheckMate 066).

Patients in the MDX010-20 trial could therefore be considered to have a poorer prognosis based on some of these factors.

The prognostic factors were included in the covariate-adjusted analysis, thus attempting to control for differences between the trials (CS Table 27). The proportion of patients with complete covariate data was high (e.g. 199 of 210 (95%) nivolumab-treated patients in CheckMate 066, CS Table 28), lessening any bias due to missing data.

The ERG considers the approach used to adjust for covariates to be generally reasonable. However, the following issues may cause uncertainty in the estimates obtained:

- The ERG notes that the Korn and colleagues[21] meta-analysis identified four significant covariates for OS, and three for PFS. The CS included a greater number (nine; see CS Table 27). It is not clear from the Korn and colleagues[21] study which prognostic factors could be applicable specifically to the outcomes analysed to inform the cost-effectiveness analysis (i.e. TTP; PrePS; PPS). Furthermore, Korn and colleagues[21] state that controlling for these prognostic variables eliminated the between-trial variability in one-year OS rates, but not in six-month PFS rates (where there was residual between-trial variation). This raises the question of whether the between-trial differences in prognostic factors were adequately adjusted for, and whether the covariates identified by Korn and colleagues[21] are applicable to the analyses of the alternative outcomes in the CS, such as TTP.
- The survival models adjusted for covariates had relatively small sample sizes for some of the time periods and outcomes considered and in many cases the prognostic factors were not significant at the 95% level (e.g. CS Table 30 and CS Table 32). In some cases treatment effects were also non-significant (e.g. ipilimumab - CS Table 32). The

non-significance of the prognostic factors may arise because of the small sample sizes, and/or the fact that they are not prognostic for the outcomes considered such as TTP (as discussed above). The ERG notes that TTP (post 100 days) was one of the most influential parameters in the CS deterministic sensitivity analysis (CS section 5.4.2).

- The CS notes (p. 113) that non-significant prognostic factors were retained in the various models in order to fully adjust for them, and to allow more flexibility within the economic model for different patient populations. This is a reasonable approach to take in this context, although many prognostic factors were adjusted for, and it is possible that they were not evenly distributed in the sample patient population – some subgroups may contain more patients than others. For example, CS Table 26 shows that there were differences between the trials for some of the prognostic characteristics (as described above), particularly for history of brain metastases. The extent to which this between-trial imbalance in prognostic factors biases the estimates is hard to gauge without access to the data used.
- The CS examined the validity of the covariate-adjusted survival models by comparing their relative treatment effect estimates with relative treatment effect of nivolumab and ipilimumab obtained an adjusted indirect comparison (CS Table 36 p. 115, and see below). Similar results were obtained, lending support to the approach used.

Adjusted indirect comparison of nivolumab and ipilimumab

The CS also reported an adjusted indirect comparison of nivolumab and ipilimumab using what it describes as a traditional approach (CS p. 115), citing the method described by Bucher and colleagues.[22] CS Table 36 reports the results of the adjusted indirect comparison for the outcomes TTP post 100 days, PPS, OS and PFS, alongside the results for these outcomes from a Weibull parametric model. A Cox proportional hazards regression was performed for the CheckMate 066 trial and for the MDX010-20 trial to obtain HRs for nivolumab versus DTIC and for ipilimumab versus gp100 (as a proxy for DTIC), respectively. The HRs were adjusted for the same covariates as used to inform the parametric survival models (described above). The primary purpose of CS Table 36 is to compare the results of two methods of indirect comparison: the adjusted approach based on the Bucher and colleagues method;[22] and the covariate adjusted parametric survival model method (used to inform the economic model). The two methods showed similar results for nivolumab and ipilimumab. The ERG notes that the Bucher and colleagues[22] method for adjusted indirect comparisons has been widely used in the health literature,[23] and in this method the comparison of the interventions of interest is adjusted by preserving the strength of randomisation. The parametric survival model-based indirect comparison appears to

preserve randomisation through inclusion of the trial as a covariate in the analyses. Both methods are therefore appropriate in this respect.

The ERG also notes that no justification is given for use of the Weibull parametric model in CS Table 36 for comparison with the adjusted indirect comparison. Use of the Gompertz model for TTP post 100 days (as used in the economic model) would have produced an HR of 0.35 compared to the HR of 0.38 for the Weibull model, which was slightly less comparable to the 0.37 HR in the adjusted indirect comparison. Likewise, use of the log-logistic HR for PPS (used in the economic model) of 0.98 instead of the HR of 0.95 from the Weibull model would have been less comparable to the HR of 0.92 in the adjusted indirect comparison. Gompertz model-based HRs might have been used throughout Table 36 instead, for example, and might not have given such a favourable comparison to the adjusted indirect figures as the Weibull model. Therefore, a justification for use of this model in the CS would have been informative.

(ii) Indirect comparison of nivolumab to BRAF inhibitors

This comparison informed the cost-effectiveness analysis for BRAF mutation-positive patients, and also comprises comparisons of treatments from trials using a common comparator. CS Table 25 describes the comparisons made and CS Figure 35 illustrates the network diagram, replicated in Figure 2 in this report. For nivolumab compared to vemurafenib, patient-level data from CheckMate 066[4] (nivolumab arm) was compared to aggregate data from the BRIM-3 trial[12] (vemurafenib arm) linked together by DTIC, which was a comparator in both trials. The ERG assumes that patient-level data from the BRIM-3 trial were not available to the company, whereas patient-level data were available for both nivolumab and ipilimumab in the BRAF mutation-negative network, since the company markets both drugs. However, the CS goes on to describe a process to create pseudo patient-level data for vemurafenib from Kaplan-Meier curves (CS P. 118, and see below).

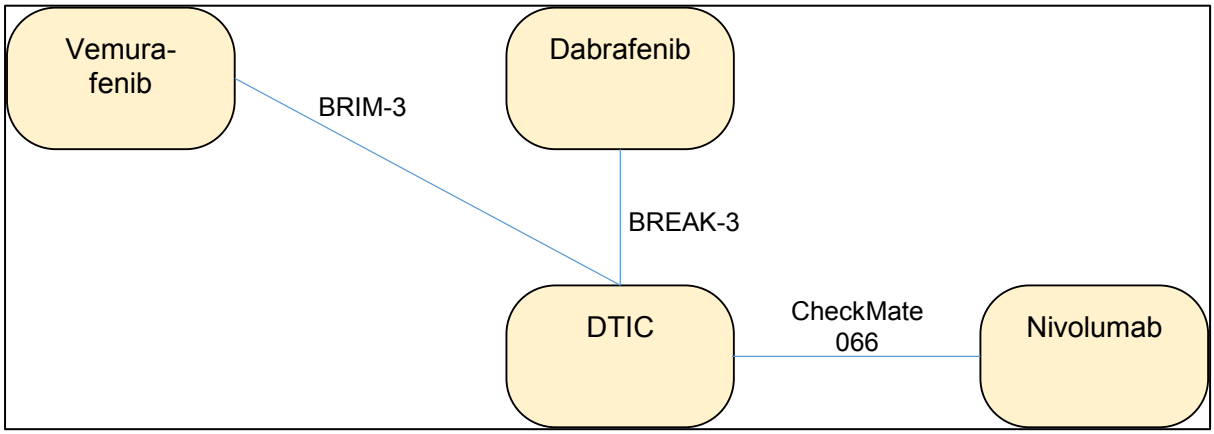

Figure 2 - Network diagram for nivolumab and BRAF inhibitor

For nivolumab compared to dabrafenib, patient-level data from CheckMate 066[4] (nivolumab arm) potentially could have been compared to aggregate data from the BREAK-3[13] trial (dabrafenib arm) linked together by DTIC (a comparator in both trials). However, the indirect comparison and survival curve fitting was subsequently restricted to nivolumab compared to vemurafenib, on the assumption that vemurafenib and dabrafenib are generally equivalent in efficacy, based on a meta-analysis used in the NICE TA321[24] of dabrafenib. (The ERG notes that the indirect comparison in the dabrafenib appraisal was not considered robust by the ERG who appraised that company submission, but that the Appraisal Committee concluded that it would not be unreasonable to assume that vemurafenib and dabrafenib have similar effect.[24])

The BRIM-3 trial[12] (vemurafenib versus DTIC) was used for the indirect comparison and survival curve fitting in preference to the BREAK-3 trial[13] (dabrafenib versus DTIC), on the basis that this was a larger trial (n=675 patients, n=250 patients, respectively) and the judgement that it was more reflective of UK patients receiving BRAF inhibitors (The journal publication for this trial[12] does not explicitly identify whether any UK patients were included, though just under two-thirds of the patients were classified as being in Western Europe, and two of the authors are affiliated with British university/hospital institutions, suggesting that there were UK centres). The CS identifies the higher lactate dehydrogenase (LDH) levels in the BRIM-3 trial[12] than the BREAK-3 trial[13] as being one factor that increased its applicability to the UK patient population, though the ERG notes that this was also higher than in the CheckMate 066[4] trial that it was indirectly compared to (CS Table 37), decreasing the similarity between these two trials. The CheckMate 066[4] trial also had a higher patient median age than the BRIM-3[12] and BREAK-3[13] trials, which potentially could confound the indirect comparison given that age is stated to be a known prognostic factor affecting treatment outcome. However, this would be accounted for in the covariate-adjusted analysis of the CheckMate 066 and BRIM-3 trials (see below).

The CS mentions a further RCT, Combi-V[25], which was included in the company's systematic review of clinical effectiveness, but subsequently not included in the indirect comparison alongside the BRIM-3 trial, as this would have necessitated multiple comparisons (stated to be necessary due to the strategy used for forming the indirect comparisons, but no further detail given, including whether it could have been used in a scenario analysis). The Combi-V trial compared vemurafenib against dabrafenib + trametinib combination therapy and does not include a DTIC arm. It is therefore not clear to the ERG

how this could have been linked to the evidence network to form a comparison with nivolumab.

Estimation of survival data and covariate-adjustment

The CS describes the process for estimating survival data from the BRIM-3 trial (CS P. 118)

- Kaplan-Meier data were estimated from the Kaplan-Meier curves for OS and PFS for vemurafenib using digitisation software.
- Using the estimated Kaplan-Meier data, pseudo patient-level data were created for vemurafenib using the Guyot 2012 method.[26] The ERG considers that this is a robust method of reconstructing survival data based upon limited published information, and it has been used in a previous technology assessment report used in a NICE appraisal.[27]
- Parametric survival curves for OS and PFS were fitted to the single-arm pseudo patient-level data – used directly in the economic model.
- The nivolumab estimates of OS and PFS (as constructed in the economic model from TTP, PrePS and PPS) were re-estimated adjusted for the observed patient characteristics from the BRIM-3 trial (CS p. 118). It is not explicitly stated which patient characteristics were included in this analysis. However, the ERG assumes it was the same covariates as used in the nivolumab versus ipilimumab comparison, based on the Korn and colleagues study[21] (CS Table 60).

Section 4.2.4 of this report describes and critiques the statistical procedures used to fit and extrapolate parametric survival curves from the trials to inform the comparisons made within the economic model.

Critical appraisal of trials included in the indirect comparison

The CS provides critical appraisal summaries for the MDX010-20 and BRIM-3 RCTs based on the Cochrane Collaboration's Risk of Bias criteria for RCTs in CS Appendix 3 (CS Appendices: Table 13 and Table 16, p. 58 and p. 61).

For MDX010-20 the ERG agrees broadly with the company's critical appraisal, with the RCT being considered at low risk of bias overall (CS Appendices Table 13, p. 58).

For BRIM-3, the ERG disagrees with the company's critical appraisal in the following aspects:

- The company concluded that randomisation was adequate. However, the ERG considers that the randomisation process (described as a minimisation procedure in

the study protocol) is unclear, and hence the risk of selection bias due to this methodological aspect is unclear.

- The company concluded that allocation concealment was adequate. However, no allocation concealment was reported for BRIM-3. The ERG therefore considers that there is a high risk of selection bias due to lack of adequate allocation concealment.
- The company concluded that although the RCT was open-label, the risk of bias would be low since there were no patient-reported outcomes specifically considered. The ERG cannot discount the possibility that outcome assessors might have introduced bias by being aware of patient allocations, e.g. when assessing and documenting disease progression, although this is unclear. The ERG therefore considers there to be an unclear risk of detection bias.

The company's overall opinion is that BRIM-3 was generally at low risk of bias except that unexpected drop-outs between groups and applicability of ITT analysis were both unclear due to the high rates of crossover permitted from DTIC to vemurafenib (i.e. unclear attrition bias risk) (CS Appendices Table 16, p. 61). The ERG concurs that risk of attrition bias is unclear, but as noted above considers that, additionally, BRIM-3 is at high risk of selection bias and unclear risk of detection bias.

Despite the above discrepancies in judgement the ERG considers that, overall, the MDX010-20 and BRIM-3 trials are appropriate for inclusion in the indirect comparison. Both trials were included as evidence considered in previous NICE melanoma appraisals.

Summary of indirect comparisons

Head-to-head comparisons of all the treatments within the scope of the STA were not conducted within the RCTs, necessitating indirect comparison. The company did not conduct a mixed treatment comparison to compare all treatments simultaneously due to clinical and methodological heterogeneity in the available evidence. As an alternative, indirect comparisons were conducted using selected RCTs from the company's systematic review of clinical effectiveness. Two separate evidence networks were created, for the comparison with ipilimumab and palliative chemotherapy (to inform the estimation of cost-effectiveness for BRAF mutation-negative patients), and for the comparison with BRAF inhibitors (for the estimation of cost-effectiveness for BRAF mutation-positive patients). Both networks used patient-level data / 'psuedo' patient-level data (BRAF mutation-negative / mutation-positive patients, respectively) from the trials to inform covariate-adjusted parametric survival models used directly in the economic model. Due to the immaturity of OS data alternative outcomes

were used: TTP and PrePS (to inform long-term extrapolations of PFS); and TTP, PrePS and PPS (to inform long-term extrapolations of OS).

The CS presents a pragmatic approach to indirectly comparing nivolumab with other treatments given the evidence limitations. The ERG considers that, overall, the approach taken is reasonable with some of the assumptions used having been accepted in previous NICE appraisals of treatments for advanced melanoma. The ERG is not aware of any relevant trials that were not included in the systematic review of clinical effectiveness, and thus absent from the indirect comparison. The trials that have been used were all multi-centre international RCTs judged to be of good methodological quality.

Trials which did not report OS were not eligible for the indirect comparison but it is not clear how many of the 44 trials identified in the systematic review would have been excluded on this criterion, and therefore how many could have provided estimates for other cancer outcomes such as PFS for potential inclusion in the analysis (though, according to the company it would not have been possible to include them unless OS events were also available – see next point).

A significant limitation is that the pivotal CheckMate 067 trial,[5] which directly compares nivolumab with ipilimumab, was not included in the indirect comparison, due to lack of available OS data. The CS does not state whether it would have been possible to have used data from the alternative outcomes (i.e. TTP, PrePS and PPS) from this trial as was done for the CheckMate 066 trial. The company clarified to the ERG that this was not possible as it requires both PFS and OS events to be available. The ERG agrees with this statement.

3.2 Summary statement of company's approach

The ERG's quality assessment of the company's systematic review of clinical effectiveness is summarised in Table 5. The processes for inclusion or exclusion of studies are described in the CS (CS p. 37-39), but the ERG notes that processes for data extraction are not described for the systematic review or the indirect comparison. Included studies were subject to critical appraisal using standard criteria recommended for use in company submissions by NICE. Overall, the ERG considers the study selection and critical appraisal processes are adequate and they appear to follow standard accepted systematic review methodology.

The ERG concludes that the submitted evidence generally reflects the decision problem defined in the CS and considers the overall risk of systematic error in the review to be low.

Table 5 - Quality assessment (Centre for Reviews and Dissemination criteria) of CS review

Quality Item: Yes/ No/ Uncertain	
1. Are any inclusion/exclusion criteria reported relating to the primary studies which address the review question?	Yes, inclusion and exclusion criteria are clearly stated.
2. Is there evidence of a substantial effort to search for all relevant research? I.e. all studies identified	Uncertain. There was substantial effort to search for all relevant published studies, and the ERG believes that all of these were identified. Ongoing trials were also searched, but these results are not provided in the CS. Only those trials are included that are expected to report data within the next 12 months.
3. Is the validity of included studies adequately assessed?	Yes. The validity of the studies is assessed in the CS using NICE-recommended criteria. However, the ERG assessment differed from the CS assessment in two criteria.
4. Is sufficient detail of the individual studies presented?	Yes, overall methodology, patient characteristics and outcomes are described in sufficient detail.
5. Are the primary studies summarised appropriately?	Yes, the primary studies are summarised appropriately, and details are presented in tables and figures. Meta-analysis was not considered possible due to heterogeneity in trials, and the ERG agrees with this.

3.3 Summary of submitted evidence

3.3.1 Summary of results for survival analysis

CheckMate 066[4], CheckMate 067[5] and CheckMate 037[9] all measure both overall survival (OS) and progression-free survival (PFS), as reported in CS Table 11 (p.49-59). The trials are still ongoing or in extended follow up, and to date OS data are only available for the CheckMate 066[4] trial. In response to the ERG's clarification question A8 the company indicated that OS data from CheckMate 067 will not be available until the number of pre-specified events (deaths) has been reached. The company does not expect this to be the case until the fourth quarter of 2016. PFS data are reported for all three trials.

The analyses demonstrate significant differences in both OS and PFS in favour of nivolumab.

Overall survival

Table 6 provides a summary of OS data from the CheckMate 066 trial (CS Table 15, p. 72).

Table 6 - Overall survival

	CheckMate 066	
	Nivolumab (n=210)	**DTIC (n=208)**
Events, n (%)	50 (23.8)	96 (46.2)
Hazard ratio (95% CI) p-value	0.42 (0.30, 0.60) <0.001	
Median OS (95% CI), months	Not reached	10.84 (9.33, 12.09)
OS rate at 6 months, % (95% CI)	84.1 (78.3, 88.5)	71.8 (64.9, 77.6)
OS rate at 12 months, % (95% CI)	72.9 (65.5, 78.9)	42.1 (33.0, 50.9)

CI = confidence interval; DTIC = dacarbazine; OS = overall survival

In this trial, OS was analysed by ITT and was based on a database lock date of 5 August 2014 (approximately 18 months after trial initiation). It contains data obtained prior to the implementation of the study protocol amendment that was made in response to a recommendation by the data monitoring committee, allowing patients who did not benefit from DTIC to cross over to nivolumab (see ERG report section 3.1.3). 'Event' is defined as death.

The OS analysis demonstrates a significant difference in deaths in favour of nivolumab. At a median follow-up of 8.9 months, a higher proportion of patients in the DTIC group had died, as compared to the nivolumab group. The corresponding hazard ratio confirms that these differences are statistically significant.

The median OS (when half of the patients have died) had not been reached in the nivolumab group at the time of the analyses, while the DTIC group had already reached a confirmed median OS, i.e. half of the patients in the DTIC group had died. The 75% OS (when a quarter of the patients have died) was reached in both the nivolumab group (10.3 months) and in the DTIC group (5.2 months) and shows an additional survival of 5.1 months in favour of nivolumab (CS narrative p. 72).

Survival rates at six months and at one year were also higher in patients randomised to nivolumab, i.e. after six months, and after one year, more patients were alive in the nivolumab group than in the DTIC group. The company comments that the one-year survival rates in the DTIC group are unusually high, potentially as a result of subsequent treatment with ipilimumab after disease progression within the first year (38% of DTIC patients).

Progression-free survival

Progression-free survival (PFS) is reported in CS Table 16 for CheckMate 066[4] (CS p. 72), and in the CS narrative for CheckMate 067[5] and CheckMate 037[9] (CS p. 75-76). Table 7 summarises PFS for these trials.

Table 7 - Progression-free survival

	CheckMate 066		CheckMate 067		CheckMate 037	
	Nivolumab (n=210)	**DTIC (n=208)**	**Nivolumab (n= 316)**	**Ipilimumab (n= 315)**	**Nivolumab (n= 122)**	**ICC (n= 60)**
Events, n (%)	108 (51.4)	163 (78.4)	174 (55.1)[b]	234 (74.3)[b]	71 (58.2)[b]	26 (43.3)[b]
Hazard ratio (95% CI) p-value	0.43 (0.34, 0.56) <0.001		0.57 (0.43, 0.76) <0.001		0.82[d]	
Median PFS (95% CI), months	5.06 (3.48, 10.81)	2.17 (2.10, 2.40)	6.9 (4.3, 9.5)[c]	2.9 (2.8, 3.4)[c]	4.67 (2.33, 6.51)	4.24 (2.14, 6.34)
PFS rate at 6 months (95% CI)	48.0 (40.8, 54.9)	18.5 (13.1, 24.6)	Not reported	Not reported	48 (38, 56)[e]	34 (18, 51)[e]
PFS rate at 12 months (95% CI)	41.8 (34.0, 49.3)	Not produced[a]	Not reported	Not reported	Not reported	Not reported

CI = confidence interval; DTIC = dacarbazine; ICC = investigator's choice chemotherapy, PFS = progression-free survival
[a] all PFS times were less than 12 months for the DTIC group.
[b] % calculated by ERG
[c] 95% CI were not reported in the CS but were taken from the trial publication[5]
[d] 95% CI and p-value are not reported. 99.99% CI is reported as 0.32-2.05.
[e] 95% CI were not reported in the CS but were taken from the trial publication[9]

In all of the trials, PFS was analysed by ITT, and 'event' was defined as death or progression. Data from all randomised patients were included in the PFS analyses for CheckMate 066[4] and CheckMate 067.[5] The PFS analysis for CheckMate 037[9] was undertaken at an interim time point, when the first 120 patients treated with nivolumab had a minimum follow-up of 6 months (median follow-up was 8.4 months). Hence, this analysis does only include a proportion of the 405 trial participants.

The PFS analyses for CheckMate 066[4] and CheckMate 067[5] demonstrate significant differences in disease progression or death between patient groups. A smaller proportion of patients in the nivolumab groups had died or experienced disease progression, as compared

to the comparator groups (DTIC or ipilimumab). The corresponding hazard ratios confirm that these differences are statistically significant.

In CheckMate 037[9] differences in PFS between patients treated with nivolumab and those treated with ICC were small. In their narrative (CS p.76), the company points out that the immaturity of the data analysed from the CheckMate 037[9] trial was primarily responsible for the uncertainty of these results, along with imbalances in prognostic factors between trial groups in favour of ICC, and high withdrawal rates in the ICC arm. The ERG agrees with the company that the observed imbalances between patient groups are likely to introduce bias. The company also states that the use of the RECIST criteria[10] for progression resulted in false-positive progression assessments in the nivolumab arm. However, the ERG notes that the RECIST criteria were also used in the assessment of patients in CheckMate 066[4] and CheckMate 067,[5] where hazard ratios for death or progression were found to be statistically significant.

3.3.2 Summary of results for response analysis

Measures of treatment response were analysed in CheckMate 066[4], CheckMate 067[5] and CheckMate 037[9] and were summarised in CS tables 17 (CS p. 77), 18 (CS p. 81), and 19 (CS p. 84). Table 8 presents a synopsis of the results from the individual analyses.

Table 8 - Response analysis

	CheckMate 066		CheckMate 067		CheckMate 037	
	Nivolumab (n=210)	**DTIC (n=208)**	**Nivolumab (n= 316)**	**Ipilimumab (n= 315)**	**Nivolumab[a] (*PP*: n= 120) (*ITT*: n=122)**	**ICC (*PP*: n= 47) (*ITT*: n=60)**
Objective response rate (ORR)						
Responders, n (%) (95% CI)	84 (40.0)[b] (33.3, 47.0)	29 (13.9)[b] (9.5, 19.4)	138 (43.7)[b] (38.1, 49.3)	60 (19.0)[b] (14.9, 23.8)	*PP*: 38 (31.7)[c] (23.5, 40.8)	5 (10.6)[c] (3.5, 23.1)
					ITT: 38 (31.1)[c] (23.1, 40.2)	5 (8.3)[c] (2.8, 18.4)
Best overall response CR, n (%) PR, n (%)	16 (7.6) 68 (32.4)	2 (1.0) 27 (13.0)	28 (8.9) 110 (34.8)	7 (2.2) 53 (16.8)	*PP*: 4 (3.3) 34 (28.3)	0 5 (10.6)
					ITT: 4 (3.3) 34 (27.9)	0 5 (8.3)
Unweighted ORR difference, % (95% CI)	26.1 (18.0, 34.1)		24.7[c]		*PP*: 21.0 (6.8, 31.7)	
					ITT: 22.8 (10.5, 32.7)	
Estimated odds ratio (95% CI) p-value	4.06 (2.52, 6.54) <0.0001		3.40 (2.02, 5.72) <0.0001		Not reported	
Duration of response						
Median (range), months	Not reached (0.0, 12.5)	5.98 (1.1, 10.0)	Not reached	Not reached	*PP*: Not reached (1.4+, 10.0+)	3.5 (1.3+, 3.5)
Time to treatment response						
Median (range), months	2.10 (1.2, 7.6)	2.10 (1.8, 3.6)	2.8 (2.3, 12.5)	2.8 (2.5, 12.4)	*PP*: 2.1 (1.6, 7.4)	3.5 (2.1, 6.1)

CI = confidence interval; CR = complete response; DTIC = dacarbazine; ITT = intention-to-treat; ORR = Objective response rate; PP = per-protocol; PR = partial response rate.
[a] CheckMate 037[9] reports both ITT and PP analyses for tumour response. [b] Confirmed response (CR+PR) as per RECIST v1.1 criteria, investigator-assessed.
[c] Confirmed response (CR+PR) as per RECIST v1.1 criteria, assessed by independent radiological review committee. [d] 95% CI not reported in the CS or in the trial publication.[5]

In CheckMate 066[4] and CheckMate 067[5], response analyses were undertaken by ITT, and data from all randomised patients were included. Tumour response was assessed by the investigators.

In CheckMate 037,[9] treatment response was assessed separately by IRRC and by investigators. IRRC-assessed response was analysed by both PP and ITT, and the ERG included these data in Table 8. PP investigator assessment is also available in the CS (CS Table 19, p. 84), but the ERG has not reported these outcomes. The ERG notes that outcome analyses in CheckMate 037 did not include all trial participants, and that analyses were undertaken at an interim time point, as described above in section *Progression-free survival*.

Overall, the analyses demonstrate significant benefit of nivolumab over comparator drugs. More patients treated with nivolumab experienced complete response (i.e. when the cancer completely disappears for a time) than those treated with alternative drugs, although the total number of patients with complete response was low in all study groups (<10%). The corresponding estimated odds ratios confirm that these differences are statistically significant.

Time to treatment response was similar between nivolumab and ipilimumab / DTIC, but appeared to be longer in the ICC group (CheckMate 037). However, the ERG notes that overall differences in time to treatment response are small.

Investigators also measured the duration of response, and treatment with nivolumab was found to be more durable than treatment with alternative drugs. The DTIC and the ICC study groups had already reached a confirmed median duration of response, i.e. half of the patients treated with DTIC or ICC were no longer experiencing benefit from treatment. In contrast, this end point was not reached in any of the nivolumab study groups and in the ipilimumab group of the CheckMate 067 trial, indicating that most patients were still experiencing treatment response at the time of analysis. The CS states that the longest duration of response observed in the nivolumab group was over 12 months at the time of analysis. A high proportion of patients continue to experience treatment response in all of the nivolumab trials and the company expects further increase in treatment duration to be found at the next data analysis (CS p. 77, p. 81, and 84).

The CS also comments on tumour burden, and changes in tumour burden are presented as waterfall plots in CS figures 15 for CheckMate 066[4] (CS p. 79), 17 for CheckMate 067,[5] and 20 for CheckMate 037[9] (CS p. 86). In all of the trials, more patients in the nivolumab groups experienced a reduction in tumour size, and achieved at least a partial response, compared with patients in the comparator groups. A best reduction in tumour size of at least 50% was reported in the majority of responding patients in the nivolumab groups of CheckMate 066[4] and CheckMate 037[9] CS p. 78 and p. 85). The median change in tumour size reported in CheckMate 067[5] was -34.5% (i.e. reduction in tumour size by more than one third) in the nivolumab group, compared to +5.9% (i.e. increase in tumour size) in the ipilimumab group (CS p. 81). Median change in tumour burden was not reported for CheckMate 066[4] and CheckMate 037.[9]

Post-RECIST criteria progression response

Continued treatment after disease progression was permitted in CheckMate 066[4] and CheckMate 067[5] for patients who experienced clinical benefit and who were tolerating the treatment. Patients in the nivolumab group of CheckMate 037[9] were also offered treatment after progression. In all three trials progression was defined by RECIST criteria (version 1.1)[10] and suitability for treatment continuation was determined by the investigators. Post-RECIST progression treatment response was reported in the CS narrative for each of the trials (CS p. 80 for CheckMate 066; CS p. 83 for CheckMate 067; CS p. 87 for CheckMate 037). In addition, the CS presents graphic representations of response patterns in Figure 16 (CS p. 80) for CheckMate 066, Figure 18 (CS p. 83) for CheckMate 067, and Figure 21 (CS p. 87) for CheckMate 037. The ERG presents a summary of these outcomes in Table 9.

Table 9 - Post RECIST progression response

	CheckMate 066[a]		CheckMate 067[b]		CheckMate 037[c]
	Nivolumab	**DTIC**	**Nivolumab**	**Ipilimumab**	**Nivolumab**
Patients treated post-progression, n	54	49	86	99	37
Responders, n (%)[d]	12 (22.2)[e]	2 (4.1)[e]	Not reported	Not reported	10 (27.0)

[a] Population described in the CS as "all treated patients" (CS p. 80).
[b] Population described in the CS as "patients with a best ORR of progressive disease" (CS p. 83).
[c] Population described in the CS as "all treated nivolumab patients at the time of interim analysis" (CS p. 87).
[d] Described in the CS as having developed or maintained a target lesion reduction of >30% compared to baseline after initial RECIST defined progression.
[e] The ERG notes that in the trial journal paper[4] the corresponding figures are 17 (31%) and 8 (16%), as discussed in section 3.1.7 of this report.

Of all patients treated with nivolumab beyond RECIST-defined progression, 22.2% in CheckMate 066 and 27.0% CheckMate 037 developed or maintained a target lesion reduction of >30% compared to baseline after progression. In comparison, only 4.1% of DTIC-treated patients experienced benefit from treatment beyond progression. As described

in Section 3.1.7 of this report, the ERG notes that the CheckMate 066 journal paper[4] reports post-progression treatment response in 17 (31%) nivolumab-treated patients and in 8 (16%) patients treated with DTIC, at odds with the figures given in the CS (and reproduced in this report in Table 9). The reason for this discrepancy is not clear. Post-progression response data were not reported for CheckMate 067, but the company states in their narrative that "many" nivolumab-treated patients experienced treatment response (CS p. 83).

The ERG concludes that a proportion of patients appear to benefit from continued nivolumab treatment beyond disease progression and the ERG would support the company's statement that treatment to progression may not always be reasonable in clinical practice (CS p.18). However, the duration of post progression treatment benefits remains unknown, as the trials are still ongoing or in extended follow-up.

3.3.3 Summary of health related quality of life

The CS provides an overview of results from CheckMate 066 for the EORTC-QLQ-C30 (which has 15 subscales), the EQ-5D utility index and the EQ-5D VAS, summarising narratively the differences in scores between the nivolumab and DTIC groups and describing changes in scores over time and in relation to baseline values (CS p. 87-89). No results for the WPAI:GH instrument are reported in the CS. Clinically meaningful differences in scores are defined in the CS by minimally important differences cited in the literature (EORTC-QLQ-C30 ≥10 points; EQ-5D utility index ≥0.08 points; EQ-5D VAS ≥7 points). The CS also presents hazard ratios (nivolumab versus DTIC; Cox proportional hazards regression models) for the time from randomisation to first decline in HRQoL and also for the time to first improvement in HRQoL, which is defined as the minimally important difference for the instrument as applied at the patient-level (CS Table 20). Due to the interim nature of the analyses, no HRQoL results from CheckMate 037 or CheckMate 067 are presented in the CS. Upon request of the ERG (clarification question A6), the company provided additional (interim) HRQoL data for CheckMate 066 (see below) and confirmed that no further data are currently available for CheckMate 037 or CheckMate 067, although partial HRQoL results from CheckMate 067 are expected in the second half of November 2015.

Initial EQ-5D and EORTC-QLQ-C30 scores as reported in the CS

The initial HRQoL results presented in the CS (p. 87-89) are from the CSR for CheckMate 066 and an abstract by Long and colleagues.[28] The CS states that the completion rates at baseline for EORTC-QLQ-C30 were 79% for the nivolumab group and 78% for the DTIC

group and adjusted completion rates (i.e. based on the numbers of patients remaining in the study) remained ≥70% up to visit week 73. Adjusted completion rates for the EQ-5D utilities index were 70% in the nivolumab group and 69% in the DTIC group and the CS states they remained similar throughout the study. However, the CS points out that due to a high attrition rate in the DTIC arm from week 13 there is high uncertainty with the HRQoL analysis after this time. No completion rates for the EQ-5D VAS are reported in the CS and the reasons for non-completion of the EORTC-QLQ-C30 and EQ-5D utility index are not specified.

The CS concludes that nivolumab does not impair HRQoL and in some cases HRQoL improved relative to baseline. However, the CS (p. 87-88) does not report individual scores for all analysis time points and it is therefore difficult to get a clear picture from the CS of whether there are any overall patterns in scores for the EORTC-QLQ-C30 and EQ-5D instruments.

Additional HRQoL data provided in the company's clarification response

The additional HRQoL data provided by the company for CheckMate 066 at the request of the ERG include graphs which clarify the time course of changes in the HRQoL measures. These graphical presentations demonstrate that EQ-5D utility index scores and EORTC QLQ-C30 global health status subscale scores were consistently higher for nivolumab than DTIC at baseline and this difference persisted throughout the study (Figures 3 and 5 in the company's clarification response). The graph for EQ-5D (Figure 3 in the company's clarification document) is reproduced in Figure 3 below.

In the clarification document the company points out that improvement in the EQ-5D utility score for nivolumab at week 37 was greater than the minimal important difference (0.08), indicating clinically meaningful improvement. However, the company does not clarify whether this difference was statistically significant. The ERG notes that uncertainty in Figures 3 to 6 in the company's clarification response (and as reproduced in Figure 3 below) is represented by standard errors rather than 95% confidence intervals; if presented instead as 95% confidence intervals there would be substantial overlap of the intervals for nivolumab and DTIC, which would indicate no statistically significant differences between the drugs for many of the time points analysed.

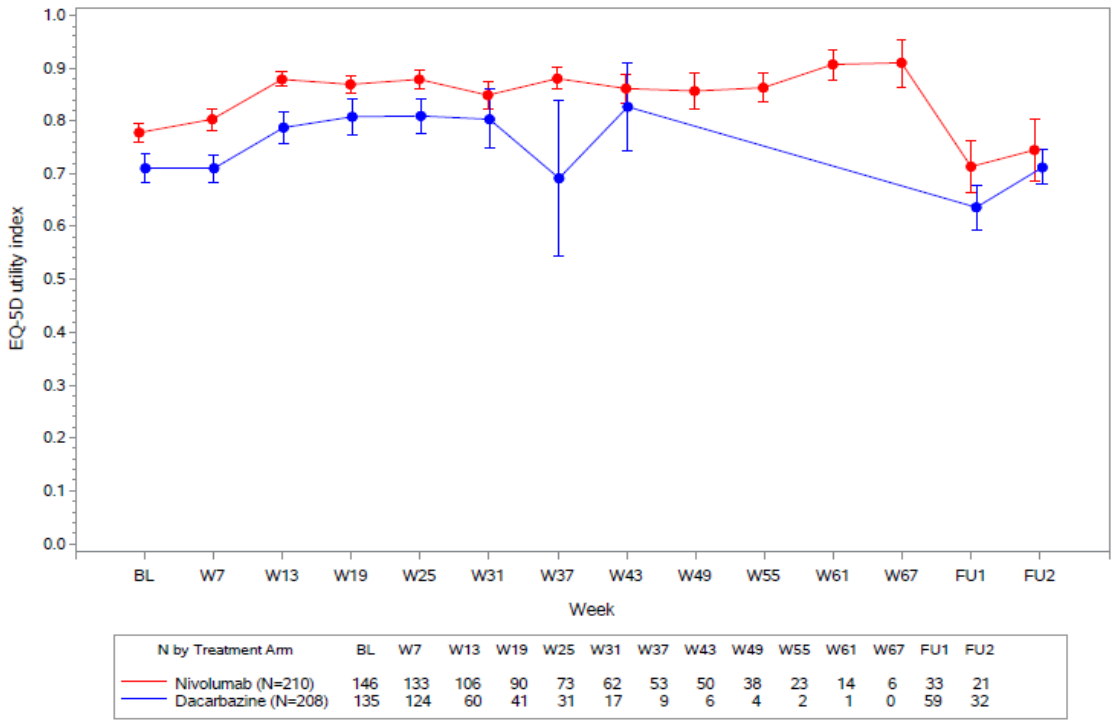

Figure 3 - Mean (±SE) EQ-5D utility index scores in the nivolumab and DTIC arms of CheckMate 066

The additional HRQoL data provided by the company also include graphs which show the change from baseline in EQ-5D utility index and EORTC QLQ-C30 global health status subscale scores (Figures 4 and 6 in the company's clarification response). These show that, when the baseline scores are taken into account, there are no consistent differences between nivolumab and DTIC and there is also no discernible improvement relative to baseline for the nivolumab arm. The graph for change from baseline in the EQ-5D utility index (Figure 4 in the company's clarification document) is reproduced in Figure 4 below. Given that the error bars presented in the graph are standard errors and error bars based on 95% confidence intervals would be wider, it appears unlikely that any of the differences between nivolumab and DTIC in Figure 4 could be considered statistically significant.

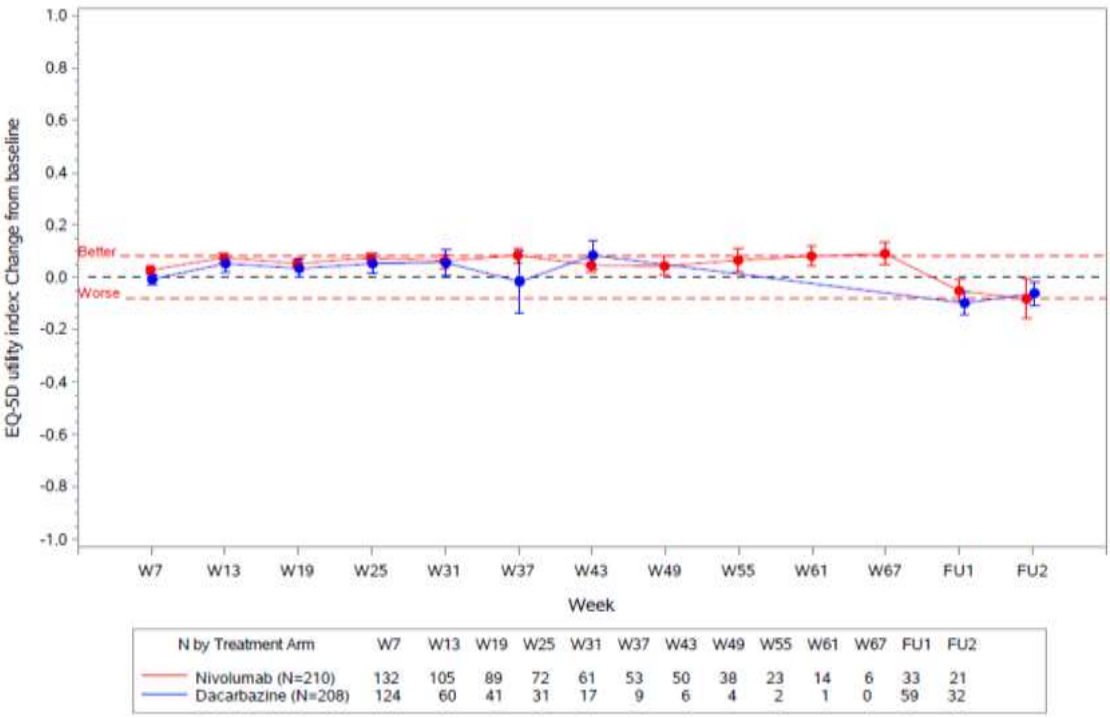

Figure 4 - Mean (±SE) changes from baseline in the EQ-5D utility index scores for the nivolumab and DTIC arms of CheckMate 066

In their clarification response, the company presents an analysis of statistically significant and/or clinically meaningful changes from baseline for nine of the EORTC QLQ-C30 subscales (Table 5 in the company's clarification response). This analysis includes 72 pairwise statistical comparisons between nivolumab and DTIC. The ERG considers that such a large number of multiple comparisons would inflate the rate of type I statistical error, potentially resulting in spurious conclusions about differences in HRQoL scores between the nivolumab and DTIC arms. Overall, the ERG's interpretation is that whilst there may be positive impacts of nivolumab on EORTC-QLQ-C30 scores relative to the baseline scores these appear to be transient and uncertain, with no clear indication of a consistent long-term improvement for any of the instrument's subscales.

Time to first decline or improvement in HRQoL as reported in the CS

The CS reports that regression analysis of time to first decline in HRQoL (CS Table 20) suggests that nivolumab had a favourable effect (HR <1.0) compared to DTIC for most of the 15 subscales of the EORTC-QLQ-C30, as well as for the EQ-5D utility index (but not for the EQ-5D VAS). Statistical significance of HR is indicated where 95% CI for the HR do not include 1.0. The largest differences in time to first decline in EORTC-QLQ-C30 subscales were for nausea and vomiting (HR=0.43 [95% CI 0.28 to 0.67]; p<0.001), dyspnoea (HR=0.50 [95% CI 0.33 to 0.75]; p<0.001), appetite loss (HR=0.43 [95% CI 0.29 to 0.65];

p<0.001), and constipation (HR=0.51 [95% CI 0.34 to 0.76]; p<0.001). Subscales of the EORTC QLQ-C30 that demonstrated no significant difference in time to first decline between nivolumab and DTIC were fatigue (HR=0.74 [95% CI 0.55 to 1.00]), diarrhoea (HR=0.87 [95% CI 0.53 to 1.43]), and financial difficulties (HR=0.66 [95% CI 0.41 to 1.05]). The time to first decline in the EQ-5D utility index favoured nivolumab (HR=0.55 [95% CI 0.38 to 0.80]; p=0.002) whereas there was no difference between nivolumab and DTIC for the time to first decline of EQ-5D VAS scores (HR=0.82 [95% CI 0.59 to 1.14]).

In contrast to the time to first decline in HRQoL, the CS provides only a brief summary of the Cox proportional hazards regression analysis results for time to first improvement in HRQoL (CS p. 88). The CS reports that time to first improvement favoured nivolumab over DTIC (i.e. HR > 1.0) for four of the 15 subscales of the EORTC-QLQ-C30. These were: global health (HR=1.52; p=0.043); physical functioning (HR=1.92; p=0.027); fatigue (HR=1.69; p=0.008); and dyspnoea (HR=2.20; p=0.013) (no 95% CI for the HR were reported). The CS also reports that time to first improvement in the EQ-5D utility index favoured nivolumab (HR=1.86; p=0.002).

Although time to first decline appears to favour nivolumab for most of the HRQoL scales assessed, including the EQ-5D utility index, the ERG notes that the method of analysis is not clearly explained in the CS, particularly with regard to whether unbalanced attrition between the trials arms after week 13 could have influenced the reported outcomes (the CS does not explicitly state which time periods are covered by the regression analyses). The ERG also notes that any initial improvements in HRQoL suggested by these Cox proportional hazards regression analyses did not appear to translate into longer-term HRQoL benefits to patients. For these reasons, and given the interim nature of the analyses, the ERG suggests that these findings should be interpreted with caution.

In summary, based on the interim HRQoL evidence presented in the CS and in the company's clarification response, the ERG agrees with the company's conclusion that nivolumab does not impair HRQoL (relative to baseline), but the ERG notes that there is no current evidence that nivolumab leads to a consistent and sustained improvement in HRQoL. Although the company's analyses suggest that nivolumab has a favourable time to first decline in HRQoL and, to a lesser extent, favourable time to first improvement in HRQoL when compared to DTIC, the best available evidence from the initial analyses does not currently suggest that this translates into longer-term HRQoL benefits.

3.3.4 Summary of results for sub-group analysis

The company undertook pre-defined subgroup analyses for most baseline characteristics (e.g. age, M stage, ECOG performance status, history of brain metastases, etc), and these are reported in the CS (p. 89-91) for CheckMate 066 (OS and response),[4] and in CS Appendix 7 for CheckMate 067(PFS; response)[5] and CheckMate 037 (response)[9] (CS Appendices p. 77-80). Forest plots are provided for all CheckMate trials.

In most subgroup analyses, outcomes were found to be in favour of nivolumab, indicating that nivolumab-treated patients benefited more compared to those treated with alternative drugs. In CheckMate 037,[9] several CIs crossed zero, indicating that nivolumab may not be effective for certain subgroups in terms of ORR (e.g. patients with BRAF mutation, ECOG PS 1, LDH above upper limit of normal, or negative PD-L1 status, among others). Subgroup analyses for CheckMate 066[4] and CheckMate 067[5] also indicated that some subgroups may not experience survival benefit from nivolumab (e.g. patients aged ≥75). However, some of these subgroups are very small and the ERG believes that the analyses should be interpreted with caution.

The ERG presents a summary of the findings from subgroup analyses by PD-L1 expression status and by BRAF mutation status below.

Subgroup analysis by PD-L1 expression status

A subgroup analysis of overall survival (OS) by PD-L1 expression status is presented for the CheckMate 066[4] trial in CS Table 21 (CS p. 90), replicated here in Table 10.

Table 10 - Overall survival by PD-L1 expression status (CheckMate 066)

	Nivolumab (n=210)	**DTIC (n=208)**
PD-L1-positive patients, n (%)	74 (35.2)	74 (35.6)
Events, n (%)	11 (14.9)	29 (39.2)
Median OS (95% CI), months	Not reached	12.39 (9.17, not reached)
Unstratified hazard ratio (95% CI)	0.30 (0.15, 0.60)	
PD-L1-negative/indeterminate patients, n (%)	136 (64.8)	134 (64.4)
Events, n (%)	39 (28.7)	67 (50.0)
Median OS (95% CI), months	Not reached	10.22 (7.59, 11.83)
Unstratified hazard ratio (95% CI)	0.48 (0.32, 0.71)	

CI = confidence interval; DTIC = dacarbazine; OS = overall survival; PD-L1 = programmed death-ligand-1.

The ERG notes that both PD-L1-positive and PD-L1-negative patients benefited from nivolumab treatment, although the proportion of patients who died was almost twice as high in the PD-L1-negative group (28.7%) than in the PD-L1-positive group (14.9%). For both PD-L1-subgroups the median OS was not reached in the nivolumab arm, i.e. more than half of these patients were still alive at the time of analyses, whereas more than half of the patients treated with DTIC had died.

A brief narrative of progression-free survival (PFS) by PD-L1 status was supplied for CheckMate 067[5] (CS Appendices p. 77). Again, patients benefited from nivolumab treatment irrespective of PD-L1 status, although median PFS in the PD-L1 positive group was longer than in the PD-L1 negative group (14.0 months compared to 5.3 months). Median PFS in PD-L1 positive ipilimumab-treated patients was 3.9 months, and 2.8 months in PD-L1 negative patients.

Objective response rates were reported by PD-L1 expression status for CheckMate 066[4] (CS Table 22, p. 90) and CheckMate 037[9] (Table 20, CS Appendices p. 79). A brief narrative ORR was also supplied for CheckMate 067[5] (described as a post-hoc analysis, CS Appendices p. 77). The results are summarised in Table 11.

Table 11 - Objective response rate by PD-L1 expression status

	CheckMate 066 ITT analysis		**CheckMate 067** Post-hoc ITT analysis		**CheckMate 037** PP objective response set IRRC assessment	
	Nivolumab n=(210)	**DTIC n=208)**	**Nivolumab (n=316)**	**Ipilimumab (n=315)**	**Nivolumab (n=120)**	**ICC (n=47)**
PD-L1-positive patients, n (%)	74 (35.2)	74 (35.6)	80 (25.3)	(75 (24)	55 (45.8)	22 (46.8)
Responders, n (%)(95% CI)	39 (52.7) (40.8, 64.3)	8 (10.8) (4.8, 20.2)	-	-	24 (43.6) (30.3, 57.7)	2 (9.1) (1.1, 29.2)
Unweighted ORR difference, % (95% CI)	-		-		34.5 (12.2, 49.2)	
ORR %	-	-	57.5	21.3	-	-
Odds ratio (59% CI)	-	-	5.03 (2.44, 10.37)		-	-
PD-L1-negative/in-determinate patients, n (%)	136 (64.8)	134 (64.4)	Not reported	Not reported	64 (53.3)	23 (48.9)
Responders, n (%)(95% CI)	45 (33.1) (25.2, 41.7)	21 (15.7) (10.0, 23.0)			13 (20.3) (11.3, 32.2)	3 (13.0) (2.8, 33.6)
Unweighted ORR difference, % (95% CI)	-		-		7.3 (-13.4, 21.5)	
ORR %	-	-	41.3%	17.8%	-	-
Odds ratio (59% CI)	-		3.25 (2.05, 5.13)		-	

CI = confidence interval; DTIC = dacarbazine; ICC = investigators choice chemotherapy; IRCC = independent radiological review committee; ORR = objective response rate; PD-L1 = programmed death-ligand-1.

In all of the trials, objective response rates were higher in nivolumab-treated patients with positive PD-L1 status than in nivolumab-treated patients with PD-L1 negative status. Both groups experienced higher response rates than patients treated with alternative drugs. However, the ERG notes that the lower bound of the 95% CI around the unweighted ORR difference between treatments in the PD-L1-negative subgroup fell below zero, indicating a potential better response for ICC treated patients in this subgroup. The trial journal publication[9] notes that these analyses, although pre-defined, were 'exploratory' and 'descriptive in nature' (p. 381) and that the patient sample sizes in some of the subgroups

were small. The ERG agrees that caution is required in the interpretation of these results for this reason.

Subgroup analysis by BRAF mutation status

Subgroup analyses by BRAF mutation status are included in the forest plots presented in CS appendix 7 (CS appendices p. 77-80). Median progression-free survival is reported for CheckMate 067[5] (CS Appendices Figure 3, p. 78), and objective response rate is reported for CheckMate 037[9] (CS Appendices Figure 4, p. 80). The ERG summarised the results in Table 12 and Table 13. No subgroup analyses by BRAF mutation status were undertaken for CheckMate 066[4] as this trial only included BRAF mutation-negative patients.

Table 12 - Progression-free survival by BRAF mutation status (CheckMate 067)

	Nivolumab (n= 316)	**Ipilimumab (n= 315)**
BRAF mutation-positive n (%)	98 (31.0)[a]	100 (31.7)[a]
Number of events n (%)	57 (58.2)[a]	66 (66.0)[a]
Median PFS (95% CI), months	5.62 (2.79, 9.46)	4.04 (2.79, 5.52)
Unstratified hazard ratio (95% CI)	0.77 (0.54, 1.09)	
BRAF mutation-negative n (%)	218 (69.0)[a]	215 (68.3)[a]
Number of events n	117 (53.7)[a]	168 (78.1)[a]
Median PFS (95% CI), months	7.98 (4.68, 12.68)	2.83 (2.76, 3.09)
Unstratified hazard ratio (95% CI)	0.50 (0.39, 0.63)	

CI = confidence interval; PFS = progression-free survival.
[a] % calculated by ERG.

The highest benefit in terms of PFS was observed in nivolumab-treated patients without BRAF mutation (BRAF wild-type), with median PFS of 7.98 months. All nivolumab-treated patients experienced longer PFS than those treated with ipilimumab, irrespective of BRAF mutation status. However, the 95% CI around the unstratified HR for BRAF mutation-positive patients crossed one, indicating no statistically significant difference between nivolumab and ipilimumab in this sub-group of patients.

Table 13 - Objective response rate by BRAF mutation status (CheckMate 037)

	Nivolumab (n=120)	ICC (n=47)
BRAF mutation-positive n (%)	26 (21.7)[a]	11 (23.4)[a]
Responders n (%)	6 (23.1)	1 (9.1)
ORR % (95% Exact CI)	23.1 (9.0, 43.06)	9.1 (0.2 41.3)
Unweighted ORR difference % (95% CI)	14.0 (-17.1, 34.4)	
BRAF mutation-negative n (%)	94 (78.3)[a]	36 (76.6)[a]
Responders n (%)	32 (34.0)[a]	4 (11.1)[a]
ORR % (95% Exact CI)	34.0 (24.6, 44.5)	11.1 (3.1, 26.1)
Unweighted ORR difference % (95% CI)	22.9 (6.2, 35.0)	

CI = confidence interval; ICC = investigator choice of chemotherapy; ORR = objective response rate.
[a] % calculated by ERG.

Nivolumab-treated patients experienced higher response rates than those treated with ICC, irrespective of BRAF mutation status. However, response rates were highest in patients with BRAF mutation-negative status. Furthermore, the lower bound of the 95% CI around the unweighted ORR difference between treatments in the BRAF mutation-positive subgroup fell below zero, indicating a potential better response for ICC treated patients in this subgroup. As described above, these subgroup analyses should be interpreted with caution due to the small sample size within each stratum.

3.3.5 Summary of adverse events

Adverse events (AE) are reported in CS section 4.2 (p. 134-145), and summaries of overall rates of AE and discontinuations due to AE are presented in CS Table 46 (CS p. 136) for CheckMate 066,[4] Table 48 (CS p. 140) for CheckMate 067[5], and Table 50 (CS p. 143) for CheckMate 037.[9] These data from the CS are replicated here in Table 14.

Table 14 - Adverse events

	CheckMate 066				CheckMate 067				CheckMate 037			
	Nivolumab (n=206)[a]		DTIC (n=205)[a]		Nivolumab (n= 313)[a]		Ipilimumab (n= 311)[a]		Nivolumab (n=268)[a]		ICC (n=102)[a]	
	Any grade	Grade 3-4	Any grade	Grade 3-4	Any grade	Grade 3-4	Any grade	Grade 3-4	Any grade	Grade 3-4	Any grade	Grade 3-4
All AEs, n (%)	192 (93.2)	70 (34.0)	194 (94.6)	78 (38.0)	311 (99.4)	136 (43.5)	308 (99.0)	173 (55.6)	255 (95.1)	92 (34.3)	95 (93.1)	44 (43.1)
TRAEs, n (%)	153 (74.3)	24 (11.7)	155 (75.6)	36 (17.6)	257 (82.1)	51 (16.3)	268 (86.2)	85 (27.3)	181 (67.5)	24 (9.0)	81 (79.4)	32 (31.4)
All SAEs, n (%)	64 (31.1)	43 (20.9)	78 (38.0)	54 (26.3)	113 (36.1)	88 (28.1)	162 (52.1)	119 (38.3)	118 (44.0)	78 (29.1)	22 (21.6)	16 (15.7)
TRSAEs, n (%)	19 (9.2)	12 (5.8)	18 (8.8)	12 (5.9)	25 (8.0)	18 (5.8)	69 (22.2)	51 (16.4)	17 (6.3)	12 (4.5)	10 (9.8)	9 (8.8)
DC due to AEs, n (%)	14 (6.8)	12 (5.8)	24 (11.7)	19 (9.3)	43 (13.7)	27 (8.6)	70 (22.5)	62 (19.9)	25 (9.3)	19 (7.1)	12 (11.8)	5 (4.9)
DC due to TRAEs, n (%)	5 (2.4)	4 (1.9)	7 (3.4)	5 (2.4)	24 (7.7)	16 (5.1)	46 (14.8)	41 (13.2)	6 (2.2)	6 (2.2)	8 (7.8)	3 (2.9)
Deaths relating to study drug, n (%)	0		0		1		1		0		0	

AEs = adverse events; DC = discontinuation; DTIC, dacarbazine; ICC = investigator's choice chemotherapy SAEs, serious adverse events; TRAEs, treatment-related adverse events; TRSAEs, treatment related serious adverse events.

[a] Patients who received at least one infusion of nivolumab or comparator drug (DTIC / ipilimumab / ICC).

Nearly all patients in these trials (>93%) experienced at least one AE of any grade, regardless of the study drug administered, with very little difference between nivolumab-treated patients and those treated with comparator drugs. In the majority of cases, AEs were treatment-related, and the proportion of nivolumab-treated patients experiencing treatment-related AEs (TRAEs) of any grade ranged from 67.5% in CheckMate 037[9] to 82.1% in CheckMate 067.[5] No major differences in the rate of TRAEs were observed between nivolumab and the comparator treatments.

Grade 3-4 AEs, TRAEs, and serious AEs (SAEs) appeared to occur less frequently in nivolumab-treated patients compared to alternative treatments, with the exception of CheckMate 037,[9] where a higher proportion of nivolumab-treated patients experienced grade 3-4 SAEs (29.1% vs. 15.7% in the ICC group) or discontinued nivolumab treatment due to grade 3-4 AEs (7.1% vs. 4.9% in the ICC group).

In all of the CheckMate trials, the proportion of patients who discontinued the study drug due to TRAEs was lower in the nivolumab groups than in the comparator groups. The ERG notes that in Checkmate 067[5] a higher proportion of patients discontinued treatment due to AEs of any grade (nivolumab: 13.7%; ipilimumab: 22.5) compared to CheckMate 066[4] (nivolumab: 6.8%; DTIC: 11.7%) and CheckMate 037[9] (nivolumab: 9.3; ICC: 11.8). In addition, discontinuation of treatment due to TRAEs occurred more frequently in Checkmate 067,[5] irrespective of treatment. The CS does not discuss between-trial differences in safety outcomes.

The most frequently reported TRAEs (reported in ≥15% of patients) in the nivolumab groups of the CheckMate trials were fatigue, pruritus, rash, diarrhoea, and nausea, as summarised in Table 15.

Table 15 - Most frequently reported TRAEs in nivolumab-treated patients[a]

	CheckMate 066 Nivolumab (n=206)	CheckMate 067 Nivolumab (n= 313)	CheckMate 037 Nivolumab (n=268)
Fatigue	19.9%	34.2%	25.0%
Pruritus	17.0%	18.8%	16.0%
Rash	15.0%	25.9%	
Diarrhoea	16.0%	19.2%	
Nausea	16.5%		

[a] Empty cells indicate that TRAE was reported in less than 15% of patients.

Hyperglycaemia, vomiting, pyrexia, pneumonitis, and infusion-related reaction were the only TRSAEs reported in more than one nivolumab-treated patient in CheckMate 066,[4] and each occurred in two patients. Hyperglycaemia was the only TRSAE reported in more than one patient in CheckMate 037,[9] and occurred in two patients. No TRSAEs were reported in more than 2% of patients in CheckMate 067.[5] None of the CheckMate trials reported TRSAEs leading to the discontinuation of nivolumab treatment in more than 1 patient (CheckMate 066, 037) or in more than 2% of patients (CheckMate 067).

One patient died from toxic effects of nivolumab (neutropenia) in CheckMate 067,[5] and there was also one treatment-related death in the ipilimumab group of this trial (cardiac arrest). CheckMate 066[4] and CheckMate 037[9] did not report any treatment-related deaths.

The CS also reports details of select AEs, which are defined as AEs "with a potential immunological cause that need frequent monitoring and potential intervention." These are categorised by organ system (endocrine, gastrointestinal, hepatic, pulmonary, renal, and skin) and are presented in CS Table 47 (CS p. 137-138) for CheckMate 066,[4] Table 49 (CS p. 140-141) for CheckMate 067,[5] and Table 51 (CS p. 144-145) for CheckMate 037.[9]

In nivolumab-treated patients, the most frequently reported select AE occurred in the skin (between 35.8% in CheckMate 037[9] and 53.4% in CheckMate 067[5]) and in the gastrointestinal categories (between 20.5% in CheckMate 037[9] and 31.6% in CheckMate 067[5]), and the least frequent categories were pulmonary and renal select AEs and hypersensitivity reactions (pulmonary: between 1.5% in CheckMate 066[4] and 3.0% in CheckMate 037;[9] renal: between 3.2% in CheckMate 067[5] and 6.7% in CheckMate 037;[9] hypersensitivity reactions between 3.0% in CheckMate 037 and 7.8% in CheckMate 066). Select AEs occurred more frequently in the nivolumab groups of CheckMate 066[4] and CheckMate 037,[9] with the exception of hypersensitivity reactions, which were more common in patients treated with ICC (CheckMate 037). In CheckMate 067,[5] select AEs were reported more frequently in ipilimumab-treated patients, and only select AEs in the endocrine, hepatic, and hypersensitivity reaction categories were reported more frequently in patients treated with nivolumab. Differences in rates of select AEs between nivolumab and ipilimumab were generally smaller than those observed between nivolumab and DTIC or ICC (apart from hypersensitivity reactions in CheckMate 066), and the ERG assumes that this is due to the fact that both nivolumab and ipilimumab belong to the

group of immune-therapies, while CheckMate 066[4] and CheckMate 037[9] compare nivolumab to chemotherapies.

The majority of select AEs were low-grade. Most were resolved with corticosteroids or other immunosuppressant medication, although median time to resolution was up to 18.4 weeks in the skin category, and up to 8 weeks in the hepatic category for nivolumab-treated patients. In CheckMate 066[4] and CheckMate 067,[5] some events of endocrine select AEs in nivolumab-, DTIC-, and ipilimumab-treated patients were controlled, but not resolved at the time of reporting (i.e. the median time to resolution had not been reached).

The ERG notes that nivolumab-treated patients had higher drug exposure than those receiving alternative treatments, as summarised in Table 16. Relative dose intensity of ≥90% was achieved by the majority of patients (84.0% to 91.3%) treated with nivolumab (33.3 to 88% in the comparator groups). This, together with the relatively low rates of treatment discontinuation due to TRAEs, indicates that nivolumab is generally better tolerated than the comparator drugs.

Table 16 - Treatment exposure

	CheckMate 066		**CheckMate 067**		**CheckMate 037**	
	Nivolumab (n=206)	**DTIC (n=205)**	**Nivolumab (n= 313)**	**Ipilimumab (n= 311)**	**Nivolumab (n=268)**	**ICC (n=102)**
Median number of doses	12	4	15	4	8	DTIC: 3 carboplatin + paclitaxel: 5
Median duration of therapy - months	6.5	2.1	6.6	3.0	5.3	2.0
% patients who received relative dose intensity ≥90%	91.3	52.2	~88	~88	84.0	DTIC: 71 carboplatin: 33.3 paclitaxel: 54.4

In summary, there was a lower incidence of high grade and serious AEs in nivolumab-treated patients compared to those treated with ipilimumab or chemotherapy, although nearly all trial participants experienced AEs (of any grade or category). AEs were typically those with potential immunological cause. Discontinuation rates due to AEs were also lower in the nivolumab

groups. Most of these were reported less often in nivolumab-treated patients than in patients treated with ipilimumab, and were generally resolved or controlled. The ERG's interpretation is that overall, nivolumab appeared to be better tolerated than the comparator drugs.

3.4 Summary

The ERG considers that the CS presents a generally unbiased estimate of the treatment effect of nivolumab for adults with advanced melanoma within the stated scope of the decision problem, although there are some exceptions and uncertainties as described below. The company's systematic review of clinical effectiveness followed standard procedures and is of good quality. The ERG is not aware of any additional relevant published trials that could be included. The three key CheckMate RCTs are well-designed and well-conducted and provide an appropriate evidence base to inform the assessment of clinical and cost-effectiveness of nivolumab. The trials show statistically significant differences in favour of nivolumab relative to alternative treatments in terms of measures of survival and treatment response, with a generally favourable safety profile.

The key uncertainties identified include:

1. All three of the key RCTs included by the company in their systematic review of clinical effectiveness are on-going with further follow-up results expected to be published in the next year. Consequently, some of the results reported in the CS are from interim time points, in some cases based on relatively small patient numbers/events, and are considered to be relatively immature due to lack of follow-up, notably for OS, one of the key outcomes that informs the assessment of cost-effectiveness in the CS. Although the duration of follow-up reported to date can be considered informative for a disease with relatively short survival time, the long-term survival benefit and benefits in terms of other relevant outcomes such as tumour response claimed by the company (CS section 4.13) cannot yet be fully substantiated.
2. The comparative efficacy of nivolumab with the comparator treatments in the scope and the decision problem is uncertain due to a lack of head-to-head data from clinical trials. Notably the CheckMate 067 trial directly compared nivolumab with ipilimumab but the results are not used to inform the company's economic model as OS data are not yet available due to insufficient follow-up, and the company stated in their response to a clarification question from the ERG that PFS data from this trial was not able to be used

to inform the model without OS events also being available (clarification question A9). This is a significant limitation of the analysis and the company has therefore made an indirect comparison of nivolumab with ipilimumab and nivolumab with the other comparators in the decision problem.

3. The indirect comparison is based upon a number of assumptions and covariate-adjusted survival data extrapolations. Some of these assumptions appear reasonable and are noted by the CS to have been accepted in previous NICE appraisals of treatments for advanced melanoma. Two of the key assumptions that influence the assessment of clinical effectiveness, and the modelling of cost effectiveness, are that previous melanoma treatment experience does not have an independent impact on treatment effect in advanced melanoma, and that there is no difference between treatment effects by BRAF mutation status. The ERG notes that there are potential limitations in the cited published pooled analysis of nivolumab studies[17] that has been used to support the assumption that BRAF mutation status does not affect outcomes in nivolumab-treated patients. Furthermore, pre-planned sub-group analyses in the CheckMate 067 and CheckMate 037 trials showed that BRAF mutation-negative patients had better outcomes (PFS and ORR, respectively) relative to comparators than BRAF mutation-positive patients, though caution is advised due to small patient sample sizes in some cases (NB. neither of these two trials directly inform the economic model).

4 ECONOMIC EVALUATION

4.1 Overview of company's economic evaluation

The company's submission to NICE includes:

i) a review of published economic evaluations of nivolumab for patients with advanced melanoma.

ii) a report of an economic evaluation undertaken for the NICE STA process. The cost effectiveness of nivolumab was compared with DTIC and ipilimumab for BRAF-mutation-negative patients and with dabrafenib, ipilimumab and vemurafenib for BRAF-mutation-positive patients with advanced melanoma.

In this section, an overview is presented of the company's submission. Further details and critique are provided in Section 4.2.

Company's review of published economic evaluations

A systematic search of the literature was conducted to identify economic evaluations of nivolumab for the treatment of melanoma. The company's search strategy for economic evaluations was adequate, though the ERG ran an update of the search to cover the nine months since the search was conducted. (See Section 3.1 of this report for the ERG critique of the search strategy).

The inclusion and exclusion criteria for the systematic review are listed in section 5.1.1 of the CS, p. 152. The inclusion criteria state that full economic evaluations of nivolumab or nivolumab in combination with ipilimumab in adults with advanced (unresectable or metastatic) melanoma would be included. The exclusion criteria state that studies before 1970 or not published in English would be excluded. One study was identified from screening 140 titles and abstracts, but was excluded during full paper screening as it was not a full economic evaluation.
No further cost effectiveness studies were identified through the ERG's update search but the ERG identified a conference abstract[6] through ad hoc searching that described a cost effectiveness analysis of nivolumab compared to ipilimumab for BRAF mutation-negative advanced melanoma in Australia. The study estimated that compared to ipilimumab over 10 years, nivolumab would lead to an improvement in survival of 1.58 years and 1.30 QALYs per person at a discounted net cost of AUD$77,119 per person and AUD$59,311 per QALY saved. The full report of this study is not yet available therefore the ERG has not been able to critically appraise this study.

Cost-effectiveness analysis methods

The *de novo* cost effectiveness presented in the CS uses a semi-Markov model to estimate the cost-effectiveness of nivolumab compared with DTIC and ipilimumab for treatment-naive BRAF-mutation-negative patients and with dabrafenib, ipilimumab and vemurafenib for treatment naive BRAF-mutation-positive patients with advanced melanoma. The model adopted a lifetime horizon of 40 years and a cycle length of one week. The model consists of three health states: pre-progression, progression and death (CS Figure 49, p155).

As mentioned earlier, the clinical effectiveness estimates of nivolumab used in the economic are based on the CheckMate 066 trial.[4] The company conducted covariate-adjusted indirect comparisons between comparators using patient-level data. These data were used to estimate TTP, PPS and PrePS outcomes, which were used to derive the transition probabilities between health states. Survival curves for TTP, PPS and PrePS are shown in CS Figure 50, p166; CS

Figure 51, p167; and CS Figure 52, p168 respectively for the BRAF mutation-negative analysis. Overall survival for the BRAF mutation-negative analysis is shown in CS Figure 56, p172 and for the BRAF mutation-positive analysis.

The analysis was conducted from the perspective of the NHS and Personal Social Services (PSS). The starting population of the model for patients in the BRAF mutation-negative analysis was based on the CheckMate 066[4] trial and the patient characteristics are shown in CS Table 59, p165. The starting population of the model for patients in the BRAF mutation-positive analysis were taken from the BRIM-3 trial[12] and are shown in CS Table 60, p173.

HRQoL was included in the model, using utility values collected from the CheckMate 066 trial.[4] HRQoL was applied according to the progression status and time to death (>=30 days before death; <30 days before death). These values were obtained from a data analysis of EQ-5D data (CS Table 67, p. 189). Disutility associated with adverse events was also included for endocrine disorders, diarrhoea (Grade 2+) and other adverse events (Grade 3+) (CS Table 65, p. 188).

Costs were included for treatments, adverse event, health state costs and end of life costs. The costs were sourced from MIMS,[29] NHS Reference costs 2013/4,[30] PSSRU 2014.[31] The unit drug costs and dosages are shown in CS Table 70, p193. The comparator treatments ipilimumab, dabrafenib and vemurafenib are subject to a patient access scheme (PAS) and have been offered to the NHS at a confidential discount. Resource use was estimated based on the MELODY study, an observational study of resource use in patients with advanced melanoma.[32] Resource use and unit costs are shown in CS Table 73, p. 196 and CS Table 74, p. 198. A covariate-adjusted time on treatment curve was used to estimate the proportion of patients on and off treatment for the nivolumab arm, with maximum treatment duration of two years assumed in the model (CS Figure 61, p. 179).

Deterministic sensitivity analyses were conducted on parameter estimates (CS p. 227-232) and additional scenario analyses were modelled (CS p. 235-240). Probabilistic sensitivity analyses (PSA) were also conducted and the input parameters are described in CS Table 79 (p. 202-204). Validation of the cost effectiveness analysis was conducted through external review by clinical experts and health economists. The CS provides a comparison between overall survival for patients treated with ipilimumab produced by the model compared to that from the clinical trials.

Cost effectiveness analysis results

Results from the economic model are presented (CS Section 5.7.1, p. 206-7) as incremental cost per QALY gained for nivolumab compared with its comparators for BRAF-mutation-negative for and BRAF mutation-positive patients. Total and incremental costs, life years gained (LYG) and QALYs were also reported, along with a breakdown of total costs. Results are presented with drug prices based on list prices and then for drug prices assuming PAS prices for the comparator treatments. Total costs are reported as commercial in confidence by the company for all treatments, in order to avoid calculation of the confidential PAS prices for ipilimumab and vemurafenib.

For BRAF-mutation-negative patients an incremental cost per QALY gained of £23,583 was reported for nivolumab versus DTIC (see Table 17). For BRAF-mutation-positive patients an incremental cost per QALY gained of £7,346 was reported for nivolumab versus ipilimumab (see Table 18).

Table 17 - Base case cost effectiveness results for BRAF mutation-negative patients (drug prices based on list price, CS Table 80)

Technology	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	ICER (£) incremental (QALYs)
DTIC	[redacted]	1.23			
Ipilimumab	[redacted]	2.64	£48,429	1.41	Excluded due to extended dominance
Nivolumab	[redacted]	4.31	£72,578	3.08	£23,583

Table 18 - Base case cost effectiveness results for BRAF mutation-positive patients (drug prices based on list price, CS Table 81)

Technology	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	ICER (£) incremental (QALYs)
Ipilimumab	[redacted]	2.44			
Nivolumab	[redacted]	4.27	£13,374	1.82	£7,346
Dabrafenib	[redacted]	1.69	£6,228	-2.57	Excluded due to dominance
Vemurafenib	[redacted]	1.70	£24,659	-2.56	Excluded due to dominance

In the deterministic sensitivity analyses of nintedanib, the results were presented in terms of net benefit with a willingness to pay threshold of £50,000 per QALY. The analyses showed that the

model results were most sensitive to the parameters defining the fitted parameter curves for TTP, PPS and long-term OS (CS Table 97-98, p. 233-240).

The CS summarises the results of the PSA stating that there is a 87% and 99% probability of nivolumab being cost-effective for BRAF-mutation-negative patients at a threshold willingness to pay of £30,000 and £50,000 per QALY gained (CS Figure 67, p. 218), and a 100% probability of nivolumab being cost effective for BRAF-mutation-positive patients for both thresholds (CS Figure 68, p. 219).

The CS states that the base case analyses show that nivolumab is a cost effective option for all patients with advanced melanoma versus all comparators at a cost-effectiveness threshold as low as £30,000 per QALY in BRAF-mutation-negative and BRAF-mutation-positive patients.

4.2 Critical appraisal of the company's submitted economic evaluation

Critical appraisal of company's submitted economic evaluation

The ERG has considered the methods applied in the economic evaluation according to the critical appraisal questions listed in Table 19, drawn from common checklists for economic evaluation methods (e.g. Drummond and colleagues[33]).

Table 19 - Critical appraisal checklist of economic evaluation

Item	Critical Appraisal	Reviewer Comment
Is there a well-defined question?	Yes	The decision problem is described in CS Table 1, p15.
Is there a clear description of alternatives?	Yes	The alternatives are listed in CS Table 56, p160
Has the correct patient group / population of interest been clearly stated?	Yes	However, analyses have been conducted for treatment naive patients but not for treatment experienced patients.
Is the correct comparator used?	Yes	However, DTIC has not been included within the analysis for BRAF mutation-positive patients. The ERG notes that pembrolizumab would now be another potential appropriate comparator but this was not included in the NICE scope.
Is the study type reasonable?	Yes	
Is the perspective of the analysis clearly stated?	Yes	Costs are considered from a National Health Service and Personal Social Services perspective. (CS Table 1, p15)
Is the perspective employed	Yes	Perspective is in accordance with the NICE framework.

appropriate?		
Is effectiveness of the intervention established?	Yes	Treatment effectiveness reported in the CheckMate 066 trial
Has a lifetime horizon been used for analysis (has a shorter horizon been justified)?	Yes	Time horizon is for 40 years (CS Table 55, p159).
Are the costs and consequences consistent with the perspective employed?	Yes	
Is differential timing considered?	Yes	Costs and health benefits discounted at 3.5% per year
Is incremental analysis performed?	Yes	Presented in CS Table 80 and 81, p206 for list price
Is sensitivity analysis undertaken and presented clearly?	Yes	Presented in CS Figure 76-77, p228-232 and scenario analyses presented in CS Table 97-98, p233-240

NICE reference case

The NICE reference case requirements have also been considered for critical appraisal of the submitted economic evaluation in Table 20.

Table 20 - NICE reference case requirements

NICE reference case requirements:	Included in submission	Comment
Decision problem: As per the scope developed by NICE	Yes	However, the analysis only includes treatment-naive patients.
Comparator: Alternative therapies routinely used in the UK NHS	Yes	Discussed in Section 4.2.3. DTIC has not been included within the analysis for BRAF mutation-positive patients.
Perspective on costs: NHS and PSS	Yes	
Perspective on outcomes: All health effects on individuals	Yes	
Type of economic evaluation: Cost effectiveness analysis	Yes	
Synthesis of evidence on outcomes: Based on a systematic review	Yes	Discussed in Section 4.2.4
Measure of health benefits: QALYs	Yes	Discussed in Section 4.2.5
Description of health states for QALY calculations: Use of a standardised and validated generic instrument	Yes	Discussed in Section 4.2.5
Method of preference elicitation for health state values: Choice based method (e.g. TTO, SG, not rating scale)	Yes	Discussed in Section 4.2.5
Source of preference data: Representative sample of the public	Yes	
Discount rate: 3.5% pa for costs and health effects	Yes	

PSS = personal social services; TTO = time trade off; SG = standard gamble

Overall, the methods applied in the economic analyses were appropriate and reported transparently. The company's economic evaluation conformed to NICE methodological guidance and generally met the NICE scope with a couple of exceptions.

4.2.1 Modelling approach / Model Structure

The company developed a *de novo* semi-Markov survival model consisting of three health states: progression-free; progressed; and death. In addition, the model incorporated two states relating to time to death (≥30 days and <30 days) for modelling utility. Costs were included according to treatment, time from initiation of therapy and proximity to death. A schematic of the model is presented in Figure 5. The model was developed in Microsoft Excel. Costs, QALYs and life years were presented as outputs of the model.

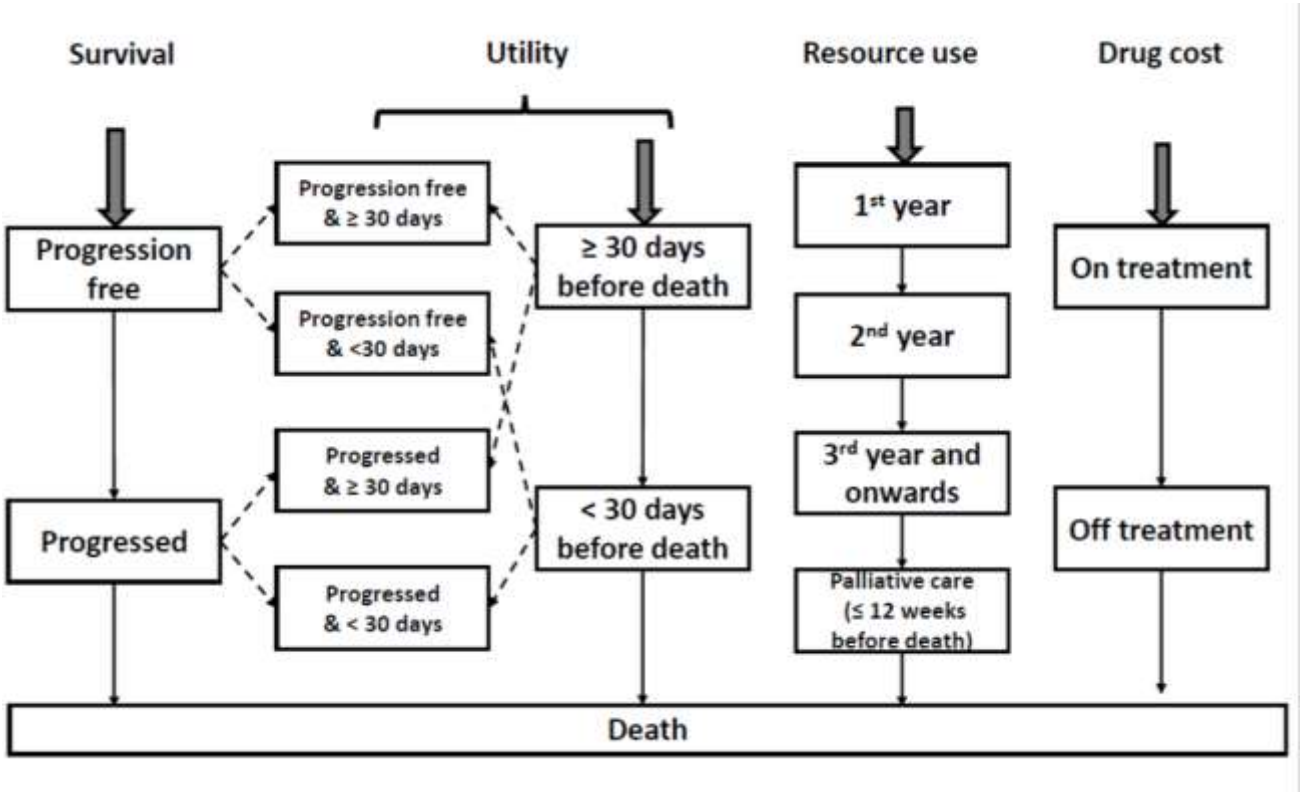

Figure 5 - A schematic of the model structure (reproduced from CS Figure 49, p 155)

The proportion of patients in each of the three health states were estimated using TTP, PPS and PrePS. Survival in the model was estimated by calculating TTP; PPS was used to estimate the time from progression to death; and PrePS was used to estimate time to death directly in instances where patients died before progression. Survival models were fit for TTP, PPS and

PrePs based on a covariate adjusted indirect comparison (described in more detail in section 3.17 and 4.2.4). The model estimated utility values based on progression-status and whether time to death was less than 30 days. The company stated that this approach was taken due to the issues arising for using RECIST criteria as a surrogate outcome for quality of life in CheckMate 066 trial.[4] The company, therefore, assigned utility values to four health states: progression-free and less than 30 days to death; progression-free and 30 or more days to death; progressed and less than 30 days to death; and progressed and 30 or more days to death. Resource use, on the other hand, was estimated based on time from treatment initiation where one-off costs were associated for treatment initiation and end of life care; and follow-up costs were used for the first year, second year, and third year after treatment initiation and for the last 12 weeks before death i.e. palliative care.

The NHS and PSS perspective was adopted in the economic model with a lifetime-horizon of 40 years and weekly cycles. This was considered as an appropriate time horizon given the median age of the patient population was 63 years for BRAF mutation-positive and 56 years for BRAF mutation-negative patients, respectively. The weekly cycle length provided sufficient time span to account for disease progression as well as treatment administration. A half-cycle correction was incorporated and costs and utilities were discounted at 3.5% p.a. as per NICE guidance.

The company cited three previous cancer technology appraisals (TA257;[34] TA258[35] and TA311[36]) as a rationale for using a state-transition method for modelling survival. Although none of these appraisals included patient groups who were treated for melanoma, the ERG agrees that state-transition modelling is a standard approach for modelling survival.

In one of the two previous NICE technology appraisals of ipilimumab (TA319),[16] a 'treatment-sequencing' approach was used to assess the cost effectiveness of ipilimumab in previously untreated melanoma in two patient groups: BRAF V600 mutation-positive (who received first-line treatment with ipilimumab, DTIC, or vemurafenib); and BRAF 600 mutation-negative patients (who received first-line treatment with ipilimumab or DTIC). The health states were defined by different lines of treatments that patients followed. The approach was criticised for some inherent inconsistencies with the evidence base, details of which are discussed elsewhere.[16] The other appraisal of ipilimumab for previously-treated melanoma patients (TA268)[20] used a 'partitioned-survival' model consisting of four health states: baseline disease, non-progressive disease, progressive disease and death. A Markov cohort model was

developed where one cohort received ipilimumab and the other cohort received best supportive care.

Broadly, the company's model structure follows a standard pattern of modelling patient transition in oncology. The disease pathway reflected the underlying clinical process of melanoma. The company validated their modelling approach with UK health economists and clinical experts. The model extrapolated short term outcomes obtained from the CheckMate 066 clinical trial to long-term outcomes by using survival models (discussed below). Overall, the ERG agrees with the company's modelling approach.

The company made a number of assumptions in relation to the model structure and model inputs on efficacy and safety, drug costs, resource use and HRQoL which are mentioned across different sections of the CS (some of these have been discussed earlier in this report). The key model assumptions are:

i. Data from trials CheckMate 066[4] and MDX010-20,[15] are used to conduct a patient-level indirect treatment comparison to obtain comparative efficacy of nivolumab, ipilimumab and DTIC based on the following assumptions (CS section 4.10, p.93, and Section 3.1.7 and Section 4.2.4 of this report):
   - DTIC and gp100 can be considered equivalent in terms of OS and PFS.
   - Line of treatment is not considered as an independent prognostic factor and is assumed not to independently affect treatment effectiveness
   - There is no difference between treatment effects by BRAF mutation status for nivolumab
   - There is equivalence of ipilimumab 3mg/kg+gp100 and ipilimumab 3mg/kg

ii. In the base case for BRAF-mutation-positive patients, the company assumed that vemurafenib had an equal efficacy to dabrafenib, based on NICE TA321[24] (CS section 5.2.2).

iii. Based on the evidence from the phase I of CheckMate 003 trial[7] and opinion of UK based clinical experts on melanoma, the company assumed that the maximum time on treatment for nivolumab was two years (CS section 5.2.2, p.156 and CS section 5.3.2).

iv. For OS, the company used pooled ipilimumab long-term data for nivolumab which showed a plateau effect in the OS for immunotherapies beginning around year three. The company also used alternative sources for long-term survival to extrapolate long-term OS for all the treatment arms. These included the use of melanoma registry data[37]

(from year two onwards for DTIC and BRAF inhibitors in the base case), long-term ipilimumab OS data[38] (from year 3 onwards for nivolumab and ipilimumab in the base case), and general UK population mortality as background mortality (CS section 5.2.2, p.156 and section 5.3.2 p.165).

Overall, the modelling approach adopted in this submission appears to be coherent. The model structure appears to be reasonable and there are no concerns regarding the techniques used with reference to the NICE methodological guidance.[39]

4.2.2 Patient Group

The characteristics of the patients in the model are based upon the patients in the CheckMate 066[4] trial for BRAF mutation-negative patients (CS Table 59, p. 165) and the vemurafenib arm of the BRIM-3[12] trial for BRAF mutation-positive patients (CS Table 60, p. 173). The patient population is consistent with the licensed indication and that population specified in the NICE scope. The ERG notes that economic analyses have only been conducted for treatment-naive patients but not for treatment-experienced patients. This is because the CheckMate 066 trial only included treatment-naive patients. The CS states that line of treatment has not been shown to independently impact treatment effect in advanced melanoma, and argue that there is no rationale for an alternative effect in the first- and subsequent-line settings. This assumption has been accepted in previous NICE appraisals of treatments for advanced melanoma. The ERG notes that it may have been possible to repeat the analysis using data from CheckMate 037 trial which included previously treated patients but this analysis was not presented.

4.2.3 Interventions and comparators

For patients with BRAF mutation-negative melanoma, nivolumab was compared to ipilimumab and DTIC. For patients with BRAF mutation-positive melanoma, nivolumab was compared to ipilimumab, vemurafenib, and dabrafenib. The comparators included within the CS economic evaluation correspond to NICE's scope, with the exception of DTIC which has not been included within the analysis for BRAF mutation-positive patients. The CS does not provide a rationale for this omission and the ERG suggests this may have been because few BRAF mutation-positive patients would be unsuitable for a BRAF inhibitor and therefore use of DTIC in this population would be rare. The ERG has conducted a scenario analysis with DTIC included as a comparator for BRAF mutation-positive patients (Section 4.3).

The CS does not include pembrolizumab, which has now received approval in advanced melanoma after disease progression with ipilimumab (NICE TA357),[3] and in patients not previously treated with ipilimumab. The ERG notes that pembrolizumab was not within the NICE scope for this appraisal.

4.2.4 Clinical Effectiveness

The following sections describe and critique the methods used to fit and extrapolate survival models to inform the economic model. For a description and critique of the indirect comparison used to estimate the comparative clinical-effectiveness and cost-effectiveness of nivolumab see Section 3.1.7 of this report.

Transition probabilities

The proportion of patients in the progression-free, progressed and death states in each Markov cycle were derived using TTP, PPS and PrePS. The transition from progression-free to progression is derived from TTP, and transition from progression-free to death from PrePS. The death rates for patients in the progression health state are derived from PPS. The parametric survival curves for TTP, PPS and PrePS were fitted based on a covariate-adjusted indirect comparison using patient-level data from trials (CheckMate 066[4] for nivolumab and DTIC and MDX010-20[15] for ipilimumab and gp100).

An advantage of using separate survival curves for TTP, PrePS and PPS is that the use of PFS as a composite endpoint is avoided. This allows the economic model to adopt a Markov-state transition approach, rather than an area under the curve partitioned survival method (CS p101-102). However to the extent that sample sizes are smaller for these endpoints than OS and PFS, the treatment effects will be estimated less precisely. Indeed, the ERG notes that for several analyses there are non-significant treatment effects, for example ipilimumab has a non-significant treatment effect at the 95% level in the Gompertz model for TTP post 100 days (CS Table 32) and the Cox proportional hazards model for PrePS (CS Table 35). Nivolumab has a borderline non-significant treatment effect in the log-logistic model for PPS (CS Table 34). The non-significance of these treatment effects may be due in part to covariate adjustment as well as a smaller sample size.

The analyses that describe the derivation of the survival curves used three types of patient-level data analyses: parametric survival modelling, Cox proportional hazards regression modelling

and Kaplan-Meier techniques. The parametric and Cox survival models were adjusted for treatment, trial and other covariates (CS Table 27). The parametric survival modelling included analyses for six different parametric distributions: exponential, Weibull, log-Normal, log-logistic, Gompertz and generalised gamma.

Survival curve modelling: BRAF mutation-negative patients

Time to progression (TTP)

As reported earlier in this report (Section 3.1.7), the TTP survival curve was modelled separately for the first 100 days, and then post 100 days. The CS comments that there is an unrealistic clustering of progression times in the studies which makes it difficult to fit meaningful parametric survival curves to these data near to the start of the curves and therefore the data were cut at Day 100 to allow a more clinically and statistically plausible shape to the progression curve. Day 100 was chosen to ensure in both studies, patients surviving from that point will have had their first tumour assessment. The ERG notes that the Kaplan-Meier curves for TTP (CS Figure 28) for nivolumab, DTIC and ipilimumab begin to diverge at around Day 100 and that by splitting this endpoint at this time the estimated treatment difference between nivolumab and the comparators is likely to be maximised. This is because an HR based on a survival model fitted to the whole time period is likely to be smaller than an HR obtained from a model fitted to data from Day 100 onwards, as treatment effects will be averaged over the entire time period.

TTP pre-100 days uses Kaplan-Meier data adjusted by a HR estimated from a Cox proportional hazards model using covariates to control for differences between trial arms and between trials. The parameters for the Cox proportional hazards model are shown in CS Table 30. Although the CS notes that proportionality of treatment effects clearly does not hold for TTP pre-100 days based on the Kaplan-Meier curves (CS Figure 29), a proportional hazard model which includes treatment effects is still used to estimate hazard ratios for the prognostic factors for this endpoint. The CS does not report if the proportional hazards assumption of this model was satisfied. The HR applied to the Kaplan-Meier data for TTP pre-100 days is 0.987 for nivolumab, 0.999 for DTIC and 0.891 for ipilimumab (values derived from company model). The ERG has examined the sensitivity of the economic model to covariate adjustment for this endpoint and found that the base case ICER for nivolumab compared to DTIC does not vary substantively when no adjustment is applied.

Standard parametric curves were fitted to the TTP post 100 days and the fitted curves are shown in CS Figure 31. Based on Akaike information criterion (AIC) and Bayesian information criterion (BIC) values, the company stated that the Gompertz distribution provided the best fit of these distributions and was deemed to be clinically plausible and in line with long-term data available for ipilimumab. The parameters for the Gompertz distribution were derived through a covariate analysis. The indirect treatment comparison effect of nivolumab vs. ipilimumab was a HR of 0.356 (95% CI 0.165, 0.771), in favour of nivolumab. The CS comments that many of the covariates individually had modest effects on the outcome and were not statistically significant but were retained to fully adjust for prognostic factors.

Deterministic sensitivity analysis showed that the economic model was sensitive to TTP post 100 days for both nivolumab and ipilimumab (CS Figures 75 and 76).

The TTP survival curves used in the company model compared to the Kaplan-Meier data for the treatment arms are shown in Figure 6.

The ERG considers the general method used to estimate the TTP survival curves to be reasonable. The Gompertz distribution was chosen for each of the treatment arms and this provides a reasonable fit for the ipilimumab treatment arm, which has the longest time follow-up, but a poorer fit for the nivolumab treatment arms. A better approach would be to use the best-fitting distribution for each treatment arm. The effect of using alternative distributions is explored by the ERG in scenario analyses (see Section 0 of this report). The ERG also notes that the numbers of patients at risk for the DTIC and nivolumab arms are small (<5%) by day 400, which makes curve fitting more uncertain.

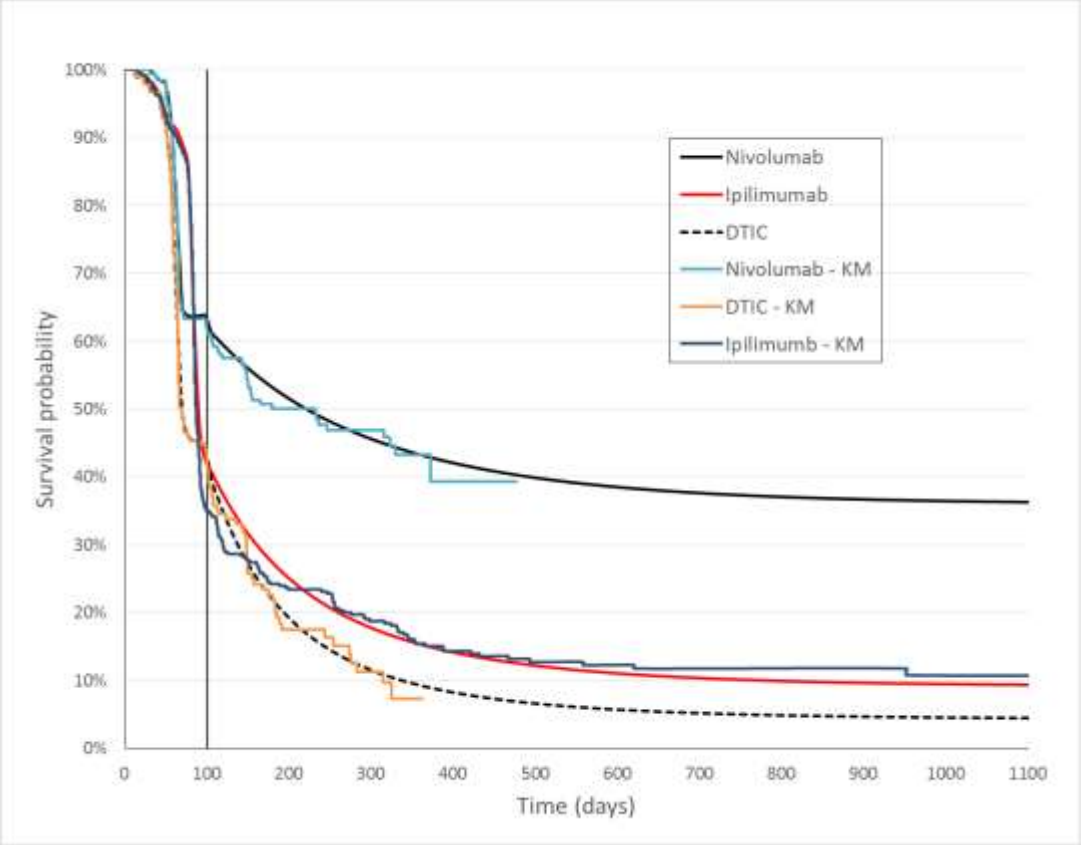

Figure 6 - Time to progression in the base case model for BRAF mutation-negative analysis over two years

Post-progression survival (PPS)

Covariate-adjusted parametric curves were also fitted to PPS for six parametric distributions. According to the AIC/BIC values, the company stated that the gamma, log-logistic and log-normal are all reasonable distributions and the company selected the log-logistic as the best-fitting/most appropriate model for use in the base case. CS Figure 51 shows the final modelled PPS for BRAF mutation-negative patients and shows PPS is similar for nivolumab and ipilimumab. The CS states that using the same model, the CS estimated the indirect treatment comparison effect of nivolumab vs ipilimumab for PSS of 0.98. The ERG notes that the choice of parametric curve has only a minor impact on the model results and considers the choice of the log-logistic distribution for PPS to be reasonable.

Pre-progression survival (PrePS)

PrePS was modelled using Kaplan-Meier data adjusted by covariates for the length of the trial follow-up. Parametric curves were fitted for PrePS, however the CS states that none of the

curves provided an acceptable visual fit to observed data. Beyond the trial follow-up duration, longer-term extrapolation was informed by the melanoma registry data,[37] long-term OS on pooled ipilimumab trials and the general population mortality. The length of the trial data varied from 477 days to 1565 days for nivolumab and ipilimumab respectively. The final modelled PrePS for BRAF mutation-negative patients is shown in CS Figure 52, p168. The CS states that the sensitivity of the economic model to assumptions around PrePS is limited due to the low number of events experienced and because the majority of the patients within the trials die following observed progression events. The ERG concurs with this statement.

Overall survival (OS)

The modelled OS for BRAF mutation-negative patients for the first three years combines the TTP, PPS, and PrePS outcomes and is presented in CS Figure 53. It indicates that, overall, the model has overall a reasonable fit to the observed data for ipilimumab and DTIC. After three years, pooled ipilimumab long-term OS[38] was used for nivolumab and ipilimumab. The CS notes that the pooled analysis showed a plateau in the OS curve beginning around year three using pooled ipilimumab trials with follow-up up to 10 years (CS Figure 55, p. 171). The long-term OS was assumed to be applicable to long-term OS for nivolumab due to similarity of mechanism of action (both are immunotherapies). Figure 7 shows the overall modelled survival for the treatment arms over a 40 year time span (CS Figure 56, p. 172). Expert clinical advice to the ERG suggested that there is some uncertainty whether nivolumab would have the same long-term plateau for OS as seen with ipilimumab. It may be that this OS plateau is unique to ipilimumab and trial evidence is not currently available for a long follow-up time period for nivolumab. The ERG tested this assumption in a scenario analysis in Section 0. OS is taken from the Melanoma registry OS by Balch and colleagues[37] for the DTIC arm from year two onwards.

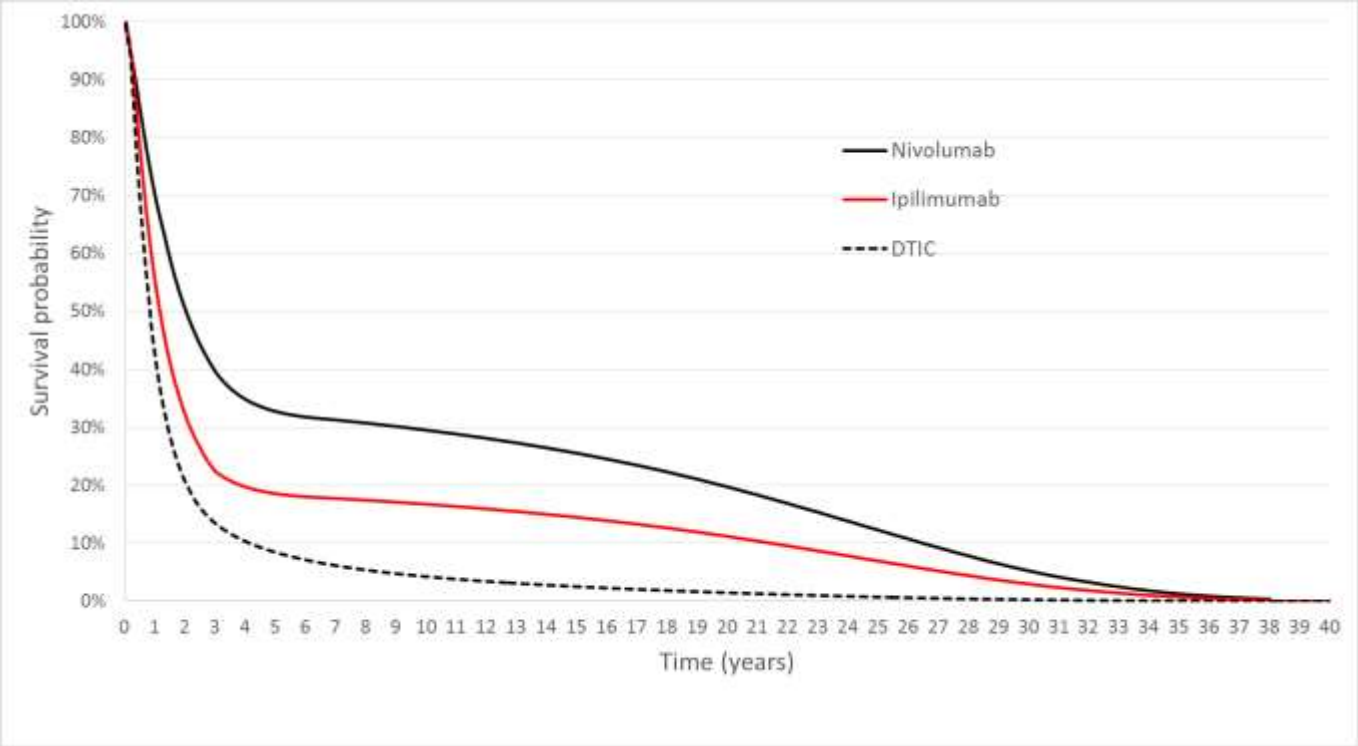

Figure 7 - Final overall survival in the base case model for the BRAF mutation-negative analysis over life time

Survival curve modelling: BRAF mutation-positive patients

The methods used for deriving transition probabilities for BRAF mutation-positive patients for nivolumab and ipilimumab-treated patients is similar to BRAF mutation-negative patients (described above), with patient characteristics in this instance based on the BRIM-3 trial.[12] As before, the survival curves are adjusted according to covariates based upon prognostic factors.

The BRAF inhibitors vemurafenib and dabrafenib were included in the analysis by fitting survival curves to Kaplan-Meier data for PFS and OS from BRIM-3[12] for vemurafenib. The company assumed that vemurafenib had an equal efficacy to dabrafenib by using a HR of 1 for OS and PFS for vemurafenib versus dabrafenib, based upon the NICE TA321[24] of dabrafenib where the Appraisal Committee determined that they have approximate equal efficacy (as discussed in Section 3.1.7 of this report).

In order to derive PFS and OS survival curves for vemurafenib, the Kaplan-Meier curves from the BRIM-3 trial were derived using digitisation software and estimating pseudo patient-level data using the Guyot 2012 method.[26] This is a method that maps from digitised curves back to Kaplan-Meier data by finding numerical solutions to the inverted Kaplan-Meier equations, using available information of events and numbers of patients at risk. As stated in Section 3.1.7 of this report, the ERG considers this to be an appropriate method to use in this circumstance. Parametric curves were then fitted to the pseudo-patient data and the log-normal distribution for OS and generalised-gamma distribution for PFS were chosen, based on the AIC/BIC values and visual fit. The proportions of patients in the model in the progression-free, progressed and dead health states were calculated directly from the PFS and OS survival curves by the area under the curve method. CS Figure 57 and CS Figure 60 show the OS and PFS in the base case model for BRAF mutation-positive analysis. The ERG notes that the costs for vemurafenib and dabrafenib are sensitive to the survival curve chosen for PFS. For example, the total costs for vemurafenib in the base case analysis was £117,655 (based on the generalised-gamma distribution), whilst using a PFS survival curve with the log-normal distribution gave total costs of £99,227. According to the AIC/BIC values, the log-normal also provided a good fit for PFS.

Time on treatment

The time spent receiving nivolumab treatment has been derived from patient-level data from the CheckMate 066 trial. Parametric curves were fitted to the data and the log-logistic curve was chosen based on the AIC/BIC scores and clinical plausibility of the distribution tail. The CS states that Gompertz curve provided the best fit but was not used in the base case because the tail of the predicted curve becomes almost horizontal from year 2 onwards and this may not be clinically plausible. The ERG notes that for TTP the company uses the Gompertz curve and considers intuitively the same curves should be used for both TTP and time on treatment. However, as noted above the ERG considers that the Gompertz should not be used for TTP. The company has provided scenario analyses using alternative distributions for time on treatment (CS Table 97, p234) which show the choice of survival curve for time on treatment has only a small effect on the model results.

The model assumes a maximum time on treatment of two years. The CS comments that treating until progression is not necessarily a realistic approach in UK clinical practice and that it would be reasonable to assume maximum treatment duration of two years in clinical practice instead. The CS reports sensitivity analyses that show that varying the maximum treatment

duration has a large effect on the model results. For example, removing the maximum treatment duration assumption increases the ICER to £68,883 per QALY. The ERG notes that the marketing authorisation for nivolumab recommends that treatment should be continued as long as clinical benefit is observed or until treatment is no longer tolerated by the patient. Expert clinical advice to the ERG suggests that clinicians and patients may be reluctant to stop treatment before disease progression, given the marketing authorisation. Expert clinical advice also suggests that patients may continue to derive benefit after they have stopped treatment and it is not clear whether patients need to be continued on treatment beyond an initial period (e.g. three months). The ERG considers that the assumption related to the treatment duration is a key issue of uncertainty in the model and investigates the effect of this uncertainty in section 0.

The dosing for ipilimumab used in the model is shown in CS Table 57 and is based on the CA184-024 trial[14] trial for doses one to four (induction 1) and MDX010-20 trial for re-induction (Induction 2 to 4). The CS notes that ipilimumab is used for a maximum of 4 doses in the UK, rather than the 16 doses used in the model.

For dabrafenib, vemurafenib and DTIC the model assumes that treatment will continue until disease progression.

Adverse events

The CS model included adverse events for endocrine disorder (any grade), diarrhoea (grade 2+) and other AEs (grade 3 +). Patient-level AE data from CheckMate 066 were used to calculate the proportion of AEs for patients in the nivolumab and DTIC arms (CS Table 61). These values differ from those reported in the trial publication and CS Table 47. The company clarified to the ERG that the values used in the model were derived from a different ad hoc analysis and that the categorisations of AEs and thresholds differed between the analyses (clarification question B5). This ad hoc analysis was done as the company's clinical advisory board felt that the reporting of the adverse events in the CSR did not capture all adverse events of relevance to clinical practice. Patient-level data for the number of hospital bed days associated with each AE from the trial was also included.

The AEs incidence for patients treated with ipilimumab is calculated as a proportion of those for nivolumab, using the ratio of adverse event rates observed in CheckMate 067. A similar method

was used to calculate the incidence of patients for dabrafenib and vemurafenib compared to DTIC using the BREAK-3 and BRIM-3[12] trials respectively.

Overall, the ERG considers that the company's approach to populate the economic model with clinical effectiveness data to be reasonable although due to the complexity of the analyses, the approach taken may appear difficult for non-statisticians to understand (and therefore suffers from a lack of accessibility and transparency) and that other, simpler, approaches may obtain similar results. As stated earlier in this report, the ERG notes that the CheckMate 067 trial data have not been used in the derivation of the survival curves due to lack of available follow-up OS data. The CheckMate 067 contains a direct comparison between nivolumab and ipilimumab and provides a key data source that has been omitted from the company's analysis. The ERG notes that there is considerable uncertainty around model results with respect to the assumptions adopted for long-term OS and time on treatment for nivolumab (explored in ERG scenario analyses – Section 4.3)

4.2.5 HRQoL

The company reports one systematic review based on the original systematic review from the NICE TA319[16] of ipilimumab, and then an update review (November 2014), for utility values and HRQoL studies for patients with advanced melanoma. The inclusion criteria specified studies reporting utilities and HRQoL data, not limited to EQ-5D.

Fifteen studies were included in the review (CS Table 64, p. 185 to 187). Thirteen studies were included from the first systematic review and two in the systematic review update. From these nine were studies directly measuring quality of life and six were cost-effectiveness studies using utilities from published articles (CS Table 64, p. 185 to 187). Details on studies found in the systematic review are provided in Appendix 13.

HRQoL was incorporated for the health states in the economic model using data from the CheckMate 066 trial[40]. Table 21, shows the mean utility values from the trial that were used to predict the utility values used within the cost effectiveness model (supplied to the ERG by the company on request, clarification question B4). The utility values, derived from EQ-5D values, defined by progression status and time to death are presented in Table 22, (CS Table 67, p. 189). Comparing these two tables it is apparent that moving from the pre-progression to post-progression states, the reduction in HRQoL observed during the trial, for both nivolumab and

DTIC is much smaller than the reduction in HRQoL predicted from the statistical model (0.03 vs 0.08).

Table 21 - Mean utility values from the CheckMate 066 trial

Mean utility by treatment arm and progression status	Utility
Nivolumab arm pre-progression	0.7892
Nivolumab arm post-progression	0.7548
DTIC arm pre-progression	0.6963
DTIC arm post-progression	0.6565

Table 22 - Quality of life (utility values) used in the company's cost effectiveness model

Health states (base case)	Mean EQ-5D utility	Range	Number of observations
Pre-progression + days left >=30 days	0.8018	Uncertainty was addressed by sampling from variance-covariance matrices assuming multivariate-normal distribution	Sample size 288 (1125 utility observations)
Pre-progression + days left >30 days	0.7795		
Post-progression + days left >=30 days	0.7277		
Post-progression + days left >30 days	0.7054		

EQ-5D data from the CheckMate 066 trial was obtained on days 1, 15, 22, and 29, continuing every six weeks for the first 12 months, and then every 12 weeks until disease progression or treatment discontinuation. For patients in the discontinued category, assessments continued every three months for the next 12 months, and then every six months thereafter.

The CS states that there were a total of 1,540 visits involving 362 patients (CS p. 182). The company conducted a statistical analysis based on this data to predict the utilities used in the model for each health state. The regression model reported was derived from a sample of 288 patients, with a total of 1125 observations. The sample size for the CheckMate 066 trial however, was a total of 418 patients for both arms and it is not clear from the CS, whether the

missing data was taken into account and how this was incorporated into the prediction model to avoid potential bias.

The utility values for health states from the regression analysis were defined by progression status and time to death and were used for all treatment arms (CS Table 67, p. 189). The CS Appendix 14 provides information regarding the statistical model used. The CS also states that utilities used in the recent ipilimumab NICE appraisal were tested in a scenario analysis.

The company reported that the final model predicts utility values using post-progression and time to death < 30 days as explanatory variables, adjusted for baseline EQ-5D values and DTIC therapy to see if there is a residual treatment effect not captured by the model. The company supplied additional information on these data upon request from the ERG that clarified that the significance cut-off used on their statistical models was 0.1.

The ERG agrees with the approach taken but expresses its reservations regarding the limited information provided on the fit of models tested. The ERG is unclear for the reason for the discrepancy between the mean trial data and the data used in the model. The ERG investigated running the model using the mean utility values from the trial, however the model results were not sensitive to changes in the utility values. An additional issue is the large amount of missing EQ-5D data, as this might have introduced bias into the estimated utility model. The ERG also notes that although the company has data for both treatment arms, they have not attempted to estimate the any differences in quality of life related to the treatments.

The impact of adverse events (AEs) on quality of life was assessed by applying a one-off utility decrement. Utility decrements for the AEs considered in the model were taken from a study by Beusterien and colleagues.[41] in which a sample of the general public evaluated outcomes for advanced melanoma in the UK and Australia.

The utility decrements for the AEs considered in the model include endocrine disorder (any grade), (disutility of -0.11), diarrhoea (Grade 2+), (disutility of -0.06) and other AEs (Grade 3+), (disutility of -0.12), (CS Table 65, p. 188). The utility decrement for other AEs associated with treatment toxicities is a mean value taken from the Beusterien and colleagues [41], consisting of a -0.11 decrement for symptomatic melanoma and -0.13 decrement for 2-5 days hospitalisation for severe toxicity. The company supplied additional information on these data upon request

from the ERG (clarification question B3). The company states that the definitions of AEs for the utility decrements had a limited match to the reported data by Beusterien and colleagues[41], so the assumptions were derived from clinical opinion received as part of the work for the company's submission of ipilimumab to NICE (TA268).[20] The assumptions used by the company are presented in Table 23**Error! Reference source not found.** in this report.

Table 23 - Assumptions used estimating utility decrements for AEs

	Model inputs	Assumptions
Endocrine disorder (any grade)	-0.11	UK decrement for 1-day in-/outpatient stay for severe toxicity (grade III/IV)
Diarrhoea (Grade 2+)	-0.06	UK decrement for Grade I/II diarrhoea
Other AEs (Grade 3+)	-0.12	Assumes 50:50 split between UK decrement for 1-day in-/outpatient stay for severe toxicity (grade III/IV) & 2–5-day hospitalisation for severe toxicity (grade III/IV)

The proportion of patients experiencing these events, and therefore the proportion of patients that these dis-utilities were applied to, were derived from CheckMate 066[40] in the nivolumab and DTIC arms. For the ipilimumab, dabrafenib and vemurafenib arms these data were estimated by deriving the proportions of patients expected to experience the adverse events in CheckMate 067[5] and applying these ratios to the BREAK-3[13] and BRIM-3[12;12] trials (CS Table 62, p. 181). These AEs' related decrements for each arm were estimated to be -0.0239 for nivolumab, -0.0325 for ipilimumab, -0.0236 for DTIC, -0.0279 for dabrafenib, and -0.0218 for vemurafenib.

The CS states that these utility decrements were applied at the start of the model, and then periodically to patients who are still on treatment, where the cycle to apply the decrement was determined by the follow-up data from the CheckMate 066[40] trial (i.e. 35 weeks). Given that the prediction model uses aggregate EQ-5D data to predict the HRQoL within each health state, an additional issue of concern, which is not clearly defined within the CS, was how the effect of AEs was marginalised avoiding double counting. The ERG notes that the treatment duration for the adverse events is based upon an annual disutility, i.e. the effect of the adverse event lasts for a year, however the company provides evidence that the adverse events last for a significantly shorter time period. The ERG is also unclear why the disutility has been applied every 35 weeks to patients. The ERG considers that the disutility has been incorrectly applied in

the model. However, the ERG notes that as the disutility is similar across all treatments, correction to the disutilities has minimal impact on model results.

Overall, the main concern for the ERG is related to the method used to incorporate HRQoL data from the trial which captures the change associated with health states but does not capture any impact of treatment on HRQoL within the health state. The ERG considers that the disutility has not been applied correctly in the economic model.

4.2.6 Resource use and costs

The main resource use and cost categories included by the company were treatment (including drug costs, cost for type of administration, one-off costs for treatment initiation and end-of-life), health state resource use such as for pre-palliative and palliative care, and resources for treating AEs.

The company conducted a systematic literature search to identify costs and resource use studies for advanced melanoma. This includes the original systematic review from the NICE TA319[16] of ipilimumab; and an update conducted up to November 2014. Overall eight studies were identified as meeting the eligibility criteria. Three of them reported only drug costs and five reported a wide range of costs and resource use. The CS however, reports that none of the studies reported on the costs or resource use associated with disease management of the newly available immunotherapies or BRAF inhibitors.

Resource use and costs for patients with advanced melanoma were included by identifying one-off resource use for treatment initiation and end of life and resource use by cycle for patients in the pre-palliative (year 1 – 3 and beyond) and palliative care period. The one-off resource use and costs for treatment initiation and end of life states were obtained from NHS Reference costs[42] and PSSRU[31] (as in NICE appraisal TA319[16] of ipilimumab).[16] While resource use and costs for patients in the pre-palliative and palliative care states were obtained from the NHS Reference case,[42] PSSRU,[31] and the Oxford Outcomes Melanoma Resource Use report.[32] Resource use data for AEs were based on patient-level CheckMate 066[40] trial data. The same sources were used to identify the unit costs for AEs, CS Table 76 (CS, p. 200) presents the unit costs and resource use for AEs. The CS reports that the unit cost data and resource use for the one-off treatment initiation and end of life costs sources used were updated according to UK clinical opinion to match current treatment practice based on responses of an advisory board

including four leading UK clinicians. They also report that these sources were used in the recent NICE appraisal TA319[16] of ipilimumab.

The resource use is modelled by dividing the patient's lifetime into health states as: first year after treatment initiation, second year, third and subsequent years following treatment initiation, and 12 weeks palliative care before death. Resource use data, proportion of patients and unit costs used in the economic model are presented in CS Table 73 for one-off resource use for treatment initiation and end of life, and CS Table 74 for cycle resource use for patients in the pre-palliative and palliative periods (CS, p. 196 to 199). The same approach as for quality of life estimates was adopted incorporating AE resource use in the model by applying this cost at the start of the model, and then periodically for patients who are still on treatment (i.e. 35 weeks).

The dosing regimen for each treatment is presented in CS Table 56 (CS, p. 160). Nivolumab is administered every two weeks by IV and the dose per administration is 236mg (i.e. 3mg/Kg, in the base case using UK patient-level weight data from the CheckMate 066 and CheckMate 067 trials, and the CA184-024 trial[14]). The recommended dosing schedule per administration for nivolumab, ipilimumab, and DTIC and the recommended daily dose for dabrafenib and vemurafenib and the drug administration costs are stated in CS Table 71 (CS, p. 194). The dose per administration and drug costs are summarised here in Table 24.

The company states that the drug unit costs of the treatments are based on the list price for nivolumab and all comparators (CS, Table 70, p. 193), with PAS discount rates explored in a scenario analysis. The list prices for drug costs have been identified from MIMS, and EMIT. The administration cost assumptions used for ipilimumab, DTIC and vemurafenib were the same as those used within the NICE TA319[16] of ipilimumab.

The administration cost was taken from NHS reference costs[42] and the treatments were assumed to be given in day care settings, every two weeks. The administration cost assumptions for ipilimumab, DTIC, and vemurafenib are the same as those within the previous ipilimumab NICE TA319[16] of ipilimumab. The active cost per administration was estimated £2,809 per infusion for nivolumab, £19,574 for ipilimumab, and £48.21 for DTIC; while, the cost per day for dabrafenib and vemurafenib was £200 and £250, respectively (Table 24). The cost for each type of administration regime was from NHS Reference Costs (2013/14).[42]

Table 24 - Dose per administration and drug costs

Drug	Dosing regimen	Dose per administration	Drug cost per administration (without PAS)	Drug cost per administration (with PAS)
Nivolumab	3mg/kg, every 2 weeks by IV	236mg	£2,809.47 per IV	n/a
Ipilimumab	3 mg/kg	236mg	£19,574.00 per IV	[redacted]
DTIC	1000mg/m2, every 3 weeks by IV	1902mg	£48.21 per IV	n/a
Dabrafenib	300mg, daily oral	300mg	£200.00 per day	[redacted]
Vemurafenib	1920mg, daily oral	1920mg	£250.00 per day	[redacted]

A one-off cost is included for BRAF inhibitors as oral chemotherapy at treatment initiation. A complete metabolic panel laboratory test cost is also added based on test requirements in the product Summary of Product Characteristics. No other additional resource use is discussed. The assumptions seem to follow the recent NICE submission TA319,[16] new assumptions are adequately described.

Overall the ERG considers the approach for costing to be reasonable. In general, the values used have been taken from standard sources and the estimates have been appropriately reported. However, there are some resource data based upon expert opinion and aspects on the adverse events and dosing information from the CheckMate 066[40] trial that the ERG is not able to check.

4.2.6 Consistency/ Model validation

The company presented a number of steps to assess the robustness of the economic model. Both health economic and clinical experts were consulted and their feedback was incorporated in the estimation of long-term survival, the treatment continuation rule for nivolumab and resource use.

The company did not report whether any checklist was used for internal validation. The company stated that the health economic and clinical experts assessed the following aspects of modelling methods and inputs (CS section 5.9, p.241-2):

Methods:

- The Markov state transition for modelling OS and PFS
- Indirect comparison of efficacy between nivolumab and ipilimumab

- Modelling time on treatment for nivolumab and the treatment discontinuation rule
- Data extrapolation beyond the trial duration and the use of external data for long-term survival

Inputs:

- Estimating utilities based on progression status and time to death
- Costs and resource use
- Safety and adverse events

Furthermore, the company verified the estimated clinical model results against those obtained from the clinical trials (CS Table 84, p. 208). For BRAF mutation-negative patients, the short term model results were compared with the trial results based on CheckMate 066[4] and results based on the BRIM-3[12] trial were used to compare the results for BRAF mutation–positive patients. Long-term model results were compared against pooled ipilimumab data. Although the short term model results for OS and PFS were comparable with those obtained from the clinical trials, the long-term model estimation of OS for ipilimumab at year 10 varied when compared to the clinical results (clinical trial results: 18%; model results: 16.8% (BRAF negative) and 13.9% (BRAF positive)). The lower OS obtained in the model estimations was explained to have resulted from small numbers of patients at risk at year 10 in the pooled analysis.

For both the patient groups, the company presented disaggregated results for both health outcomes (including QALY gains and life year gains) as well as for costs by the health states (CS Tables 85-92 p.212-7).

The company did not describe any basic input and output verification checks of the model. The ERG conducted a list of extreme value checks of the model inputs and their expected outputs to examine if the model was coded correctly. Setting the resources used to zero resulted in no costs as expected, as did using zero values for all the types of costs. Similarly, using zero value for all health state utilities and adverse events disutilities resulted in no health outcomes whereas using the value of one matched the QALYs to life years. Total costs and QALYs decreased with an increase in the discount rates which is logical and consistent. The ERG did not detect any input errors and the model calculations appeared to function correctly. Although a minor error was identified in reporting incremental costs, incremental life years gained and incremental QALYs for PAS base case results in BRAF mutation-positive patients (CS Table 83, p. 207), the model however appeared to estimate the results correctly. The logical flow of the

model appeared to work as intended and no errors were found. Overall, the model is clearly laid out, well presented, easy to navigate through and on re-running, produced results as expected.

The company has compared the modelled OS and PFS survival curves to those observed in the CheckMate 066 trial. CS Figure 53 compares the economic model predictions of OS in the base case for the first three years with observed data. It indicates that the model has overall a reasonable fit to the observed data for ipilimumab and DTIC. CS Figure 59 indicates that there is a very good fit between observed Kaplan-Meier data and PFS from the economic model for both DTIC and nivolumab (CS p. 177).

In the company's clarification response to the ERG (clarification question B2), the company has provided an analysis using the CheckMate 066 trial only, without adjusting survival curves for covariates. The analyses show similar results to those presented in the base case analysis with an ICER of £28,583 per QALY for nivolumab versus DTIC.

For external validation, the company reported that they had compared the PAS based cost-effectiveness results of the current submission with the results obtained from the PAS based analyses of the previous NICE appraisal TA319[16] of ipilimumab. Although the results of the current submission were stated to be comparable with the results in TA319, the ERG could not check this due to the commercial-in-confidential nature of the PAS price discount for ipilimumab in NICE TA319. The ERG, however, cross-checked the list price base case results of the two submissions and found significant discrepancies in the two sets of results in the submissions as shown in Table 25 and Table 26. In the BRAF mutation-negative patients, the ICERs (ipilimumab vs DTIC) obtained in TA319 were lower than that obtained by the company in the current submission. However, in the BRAF mutation-positive patients, both the analyses found vemurafenib to be dominated when compared against ipilimumab.

4.2.7 Assessment of uncertainty

The company conducted a range of sensitivity analyses including one-way sensitivity analyses, scenario analyses and PSA. Structural uncertainty was tested in the scenario analyses and heterogeneity was dealt with to some extent by running the model separately for the two patient groups: BRAF mutation-positive and BRAF mutation-negative.

Table 25 - Comparison of results obtained in the CS with TA319 for BRAF mutation-negative patients

Company submission: BRAF mutation-negative					
Treatment	**Costs**	**QALYs**	**Incremental Costs**	**Incremental QALYs**	**ICER (vs DTIC)**
DTIC	[redacted]	1.23			
Ipilimumab	[redacted]	2.64	£48,429	1.41	£34,261
Nivolumab	[redacted]	4.31	£72,578	3.08	£23,583
TA319: BRAF mutation-negative (based on CA 184-024 data for ipilimumab)					
Treatment	**Costs**	**QALYs**	**Incremental Costs**	**Incremental QALYs**	**ICER (Ipilimumab vs comparator)**
Ipilimumab	[redacted]	2.35			
DTIC	[redacted]	1.56	£13,493	0.80	£16,957
TA 319: BRAF mutation-negative (based on pooled chemotherapy naive data for ipilimumab)					
Treatment	**Costs**	**QALYs**	**Incremental Costs**	**Incremental QALYs**	**ICER (Ipilimumab vs comparator)**
Ipilimumab	[redacted]	2.50			
DTIC	[redacted]	1.55	£16,948	0.95	£17,866

Table 26 - Comparison of results obtained in the CS with TA319 for BRAF mutation-positive patients

Company submission: BRAF mutation-positive					
Treatment	**Costs**	**QALYs**	**Incremental Costs**	**Incremental QALYs**	**ICER (vs ipilimumab)**
Ipilimumab	[redacted]	2.44			
Nivolumab	[redacted]	4.27	£13,374	1.82	£7,346
Dabrafenib	[redacted]	1.69	£19,602	-2.57	Dominated
Vemurafenib	[redacted]	1.70	£38,033	-2.56	Dominated
TA 319: BRAF mutation-positive (based on CA 184-024 data for ipilimumab)					
Treatment	**Costs**	**QALYs**	**Incremental Costs**	**Incremental QALYs**	**ICER (Ipilimumab vs comparator)**
Ipilimumab	[redacted]	2.31			
DTIC	[redacted]	1.56	£23,766	0.75	£31,558
Vemurafenib	[redacted]	2.13	-£12,625	0.18	Dominated
TA319: BRAF mutation-positive (based on pooled chemotherapy naive data for ipilimumab)					
Treatment	**Costs**	**QALYs**	**Incremental Costs**	**Incremental QALYs**	**ICER (Ipilimumab vs comparator)**

Ipilimumab	[redacted]	2.45			
DTIC	[redacted]	1.55	£25,525	0.90	£28,465
Vemurafenib	[redacted]	2.10	-£9,814	0.35	Dominated

Across most of the scenario and sensitivity analyses, nivolumab was cost-effective compared to ipilimumab and DTIC at list price for BRAF-mutation-negative patients. Similarly, in BRAF-mutation-positive patients, nivolumab represented a cost-effective option when compared to ipilimumab, dabrafenib and vemurafenib at the list price for a majority of the analyses conducted. The exception to this was for scenarios related to treatment discontinuation in the nivolumab arm. [redacted]

One-way sensitivity analyses

The company conducted 53 one-way sensitivity analyses for BRAF mutation-negative patients and 58 analyses for BRAF mutation-positive patients. The parameters were varied between their upper and lower 95% confidence intervals bounds. A variation of 20% around the mean was assumed for parameters with no distribution, as stated in the model. Details of the parameters included in the analyses are listed within the model. These can be grouped under the following categories:

- Administration costs
- Patient dosing parameters including weight and height
- Proportions of patients receiving different doses of ipilimumab
- AE costs
- Resource use and costs
- Utilities (which included utility coefficients for the regression equation)
- Treatment duration
- Efficacy parameters

The results of the sensitivity analyses are presented as tornado diagrams that illustrated the 20 most influential parameters for both the patient groups (CS Figure 75 and Figure 76, p. 228-232) with the effect expressed in terms of incremental net benefit. The figures for the PAS based analyses are presented in CS Appendix 16. The model has the built-in flexibility to select up to 50 parameters in the tornado diagrams. The ICER values for the analyses are also presented in the model. The company used a willingness to pay threshold of £50,000 per QALY

for the estimation of incremental net benefit on the basis that nivolumab qualifies as an end-of-life treatment (see Section 5 of this report for discussion of end of life criteria).

Based on the company's findings, the economic model was found to be most sensitive to the parameters that defined the fitted parametric curves for TTP, PPS, long-term OS, OS/PFS for vemurafenib and time on treatment, as well as utility parameters and administration cost. The ERG considers the company's conclusions relating to the influential parameters impacting the base case results to be reasonable.

Scenario Analysis

The company also included a range of scenario analyses to assess the robustness of the model with respect to the following structural assumptions:

- fitting alternative parametric curves to TTP, PPS, long-term survival and time on treatment curve for nivolumab
- alternative approach for indirect comparison trial evidence (comparing the CheckMate 066[4] trial with the CA184-024 trial[14], rather than the MDX010-20 trial[15]) and PPS data (based on combined PPS for nivolumab and ipilimumab).
- treatment discontinuation and maximum length of treatment duration
- alternative approach to modelling dosing, drug cost and utilities
- time horizon
- discount rates

The results of the analyses are tabulated in CS Table 97 p. 233-236 (for BRAF mutation-negative) and CS Table 98 p. 237-240 (for BRAF mutation-positive) for the list price and PAS price. The findings of the company's analyses indicate that a majority of the scenarios tested did not influence the base case results and nivolumab remained cost-effective compared to the comparator drugs in both the patient groups. The exceptions were scenarios examining the effect of changing the proportion of patients continuing treatment at two years: the number of years of maximum treatment duration for nivolumab, and reducing the model time horizon to 10 years.

The scenarios relating to treatment discontinuation and maximum length of treatment discontinuation for nivolumab have the most impact on model results. The marketing authorisation for nivolumab states that treatment should be continued as long as clinical benefit is observed or until treatment is no longer tolerated by the patient (CS p. 158). The company

considered treating patients until disease progression occurs as an unrealistic approach, as a result of which they assumed that the maximum treatment duration for nivolumab was two years. The treatment duration was altered to three, four, and five years and to no maximum duration in the company scenario analyses, with ICERs increasing according to increased treatment duration.

The ERG has included additional scenario analyses to test some of the key assumptions and input parameters associated with uncertainty in the economic model (Section 4.3).

Probabilistic Sensitivity Analysis

The company performed PSA for 1000 simulations and presented the results using scatter plots (CS Figure 71-74, p. 221-224) and cost-effectiveness acceptability curves (CS Figure 67-70, p.218-220). The list of the input parameters for the analysis is presented in CS Table 79, p.202-204 and within the model. The company used normal distribution for resource use and costs; beta distribution for patient dosing parameters, proportion of patients receiving different doses of ipilimumab, and adverse event disutilities; and multivariate normal distributions for the coefficients of the regression analysis for utility. The ERG considered the assigned distributions to be appropriate; although the use of the gamma distribution for costs would have been preferable. The ERG re-ran the PSA using 1000 simulations which took approximately 10 minutes to run.

Based on the PSA results, the company concluded that in BRAF mutation-negative patients, the probabilities of nivolumab being cost effective at willingness to pay thresholds of £30,000 and £50,000 are 87% and 99% respectively for the list price and the probabilities for PAS analyses are [redacted] respectively. In the BRAF mutation-negative patients, the probabilities of the drug being cost effective at these thresholds are 100% and 100% respectively at the list price and at the PAS price the probabilities are [redacted] respectively.

The PSA results are found to be similar to those obtained in the deterministic analysis (CS Table 93-96, p.225-226).

Overall, the ERG considers that the company included a reasonable list of parameters in the PSA and the distributions used for the model parameters were appropriate.

4.2.8 Comment on validity of results with reference to methodology used

The structure of the economic model was appropriate, comprehensive and reflected the clinical pathway for patients with advanced melanoma. The economic model, developed in Microsoft Excel, was well-structured and provided the relevant data sources in a transparent way. Furthermore, the model provided graphs to enable comparison between the model results and the trial data which aided validation. The ERG did not find any errors in the coding of the model structure.

The methods chosen for the analysis were generally appropriate and conformed to NICE methodological guidelines. In general, the methods chosen to derive the survival curves were more complex than traditional methods and the complexity of the analyses may appear difficult for non-statisticians to understand and therefore limit accessibility and transparency.

The ERG identified several areas where choice of parameter was not sufficiently justified or uncertainty was not insufficiently explored. Where these concerns were identified, the ERG has conducted additional analyses, where possible, to address the uncertainty surrounding these parameters.

The ERG observed that the CheckMate 067 trial data have not been used in the derivation of the survival curves due to lack of available follow-up OS data. The CheckMate 067 contains a direct comparison between nivolumab and ipilimumab and would provide a better estimate of the comparison between these treatments that using an indirect comparison.

The ERG noted that DTIC has not been compared as a comparator in the analysis of BRAF mutation-positive patients, although it was within the NICE scope.

The ERG had reservations of the choice of survival curve used in the model for TTP for nivolumab. The model uses the Gompertz survival for all treatment comparators but the ERG suggests that other survival curves may be plausible for nivolumab.

The model assumes that the long-term survival of patients treated with nivolumab would follow a similar pattern to ipilimumab, i.e. beyond two years most patients remain alive, however there is uncertainty at present, from the trial data, whether this would indeed be the case.

For BRAF positive patients, the cost of the BRAF inhibitors was sensitive to the type of survival curve chosen for the BRAF inhibitors. The ERG noted that other survival curves are plausible that give more favourable results for the BRAF inhibitors.

4.3 Additional work undertaken by the ERG

The ERG observed a number of issues and uncertainties in the CS which are explored in this section. The additional work undertaken by the ERG is based around the following aspects:

1. Type of survival model chosen for treatment efficacy
    i. Time to progression: using the Weibull, lognormal, log-logistic and generalised gamma distributions for nivolumab patients and Gompertz distribution for DTIC and ipilimumab
    ii. Progression-free survival: using the exponential, Gompertz, log-logistic, lognormal and Weibull distributions for BRAF inhibitors (vemurafenib assumed to be same as dabrafenib)
2. Modelling method: using the data extrapolation method to model long-term survival for nivolumab
3. Using DTIC as a comparator in BRAF mutation-positive patients
4. Presentation of the ERG's preferred scenarios which includes a combination of scenarios (1), (2) and (4) outlined above and between two years and no maximum treatment duration for nivolumab.

4.3.1 Modelling TTP for nivolumab patients with the Weibull, lognormal, log-logistic and generalised gamma and Gompertz distribution for DTIC and ipilimumab

As discussed in Section 4.2.4 of this report, the ERG observed that although the Gompertz distribution provided a reasonable fit for modelling TTP for the ipilimumab arm a preferable approach to model this parameter would be to use the best fitting distribution for each treatment arm. The ERG therefore fitted the Weibull, lognormal, log-logistic and generalised gamma distributions for nivolumab, without changing the company's assigned Gompertz distribution for the DTIC and ipilimumab arms. The results are presented in Table 27 and Table 28 for BRAF mutation-negative and BRAF mutation-positive patients respectively.

Changing the survival model for nivolumab had a small impact on the incremental costs and QALYs in BRAF mutation-negative patients, increasing the ICERs (nivolumab vs DTIC) marginally, ranging from £26,483 to £27,027 from the base case ICER of £23,583. In BRAF

mutation-positive patients, the impact was similar with ICERs (nivolumab vs ipilimumab) ranging from £8,836 to £9,144, deviating from the base case ICER of £7,346.

Table 27 - Using the Weibull, log-normal, log-logistic and generalised distributions for the nivolumab arm to model time to progression at list price (BRAF mutation-negative patients)

Treatment	Distribution	Incremental Costs (vs DTIC)	Incremental QALYs (vs DTIC)	ICER (vs DTIC)	ICER (vs ipilimumab)
Nivolumab	Base case[1]	£72,578	3.08	£23,583	£14,513
Nivolumab	Weibull	£72,237	2.73	£26,483	£18,117
Nivolumab	Lognomal	£72,085	2.67	£27,027	£18,874
Nivolumab	log-logistic	£72,137	2.69	£26,829	£18,594
Nivolumab	generalised gamma	£72,098	2.67	£26,980	£18,806

[1]:Gompertz

Table 28 - Using Weibull, log-normal, log-logistic and generalised gamma distributions for the nivolumab arm to model time to progression at list price (BRAF mutation-positive patients)

Treatment	Distribution	Incremental Costs (vs ipilimumab)	Incremental QALYs (vs ipilimumab)	ICER (vs ipilimumab)
Nivolumab	Base case[1]	£13,374	1.83	£7,346
Nivolumab	Weibull	£13,060	1.48	£8,836
Nivolumab	Lognomal	£12,890	1.41	£9,144
Nivolumab	log-logistic	£12,947	1.43	£9,025
Nivolumab	generalised gamma	£12,903	1.41	£9,120
Nivolumab dominates dabrafenib and vemurafenib for all analyses				

[1]: Gompertz

4.3.2 Modelling progression-free survival using a range of distributions for BRAF inhibitors

For PFS, it was observed that the type of survival curve chosen for the BRAF inhibitors influenced the costs associated with the treatment arms in BRAF mutation-positive patients. The ERG explored this further by assigning a range of distributions (exponential, Gompertz, log-logistic, log-normal and Weibull) to the PFS in the BRAF inhibitors. Assigning different distributions influenced the total costs for both dabrafenib and vemurafenib but total QALYs in both the treatment arms remained similar to the base case values as shown in Table 29. As in the base case, the ICERs for both the BRAF inhibitors (vs ipilimumab) remained dominated for the scenarios with different survival distributions.

Table 29 - Using a range of distributions to model PFS for the BRAF inhibitors at the list price (BRAF mutation-positive patients)

Treatment		Incremental Costs (vs ipilimumab)	Incremental QALYs (vs ipilimumab)	ICER (vs nivolumab)
Dabrafenib	base case[1]	£19,602	-0.75	Dominated
Dabrafenib	exponential	£5,950	-0.75	Dominated
Dabrafenib	Gompertz	£1,783	-0.76	Dominated
Dabrafenib	log-logistic	£37,002	-0.74	Dominated
Dabrafenib	log normal	£4,860	-0.75	Dominated
Dabrafenib	Weibull	£1,538	-0.76	Dominated
Vemurafenib	base case[1]	£38,033	-0.74	Dominated
Vemurafenib	exponential	£20,964	-0.74	Dominated
Vemurafenib	Gompertz	£15,757	-0.75	Dominated
Vemurafenib	log-logistic	£59,778	-0.73	Dominated
Vemurafenib	lognormal	£19,605	-0.74	Dominated
Vemurafenib	Weibull	£15,452	-0.75	Dominated

4.3.3 Using the data extrapolation method to model long-term survival for nivolumab

The company's long-term OS analysis, based on pooled ipilimumab data showed a plateau effect for ipilimumab beginning around year three. The company assumed the same effect for the nivolumab arm in their analyses (for details, see Section 4.2.4 of this report) which might not reflect the clinical trajectory of overall survival for the nivolumab treatment arm. The ERG explored the impact of using extrapolated long-term survival data for the nivolumab arm (using the Gompertz survival curve), rather than using the pooled ipilimumab data. As shown in Table 30 and Table 31, changing the modelling method reduced the total costs of nivolumab by approximately £2,000 and reduces the QALYs gained for nivolumab in both the patient groups from the base case values.

Table 30 - Using the data extrapolation method to model long-term survival for nivolumab at the list price (BRAF mutation-negative patients)

Treatment	Costs	QALYs	Incremental Costs (vs DTIC)	Incremental QALYs (vs DTIC)	ICER (£/QALY)
DTIC	[redacted]	1.23			
Ipilimumab	[redacted]	2.64	£48,429	1.41	£34,261
Nivolumab	[redacted]	3.25	£70,761	2.02	£36,072

The ICERs for nivolumab vs DTIC increased from £23,583 in the base-case to £36,072 in BRAF mutation-negative patients. The ICERs for nivolumab compared to ipilimumab increased from £7,346 in the base case to £27,171 in BRAF mutation-positive patients.

Table 31 - Using data extrapolation method to model long-term survival for nivolumab at the list price (BRAF mutation-positive patients)

Treatment	Costs	QALYs	Incremental Costs (vs ipilimumab)	Incremental QALYs (vs ipilimumab)	ICER (£/QALY)
Ipilimumab	[redacted]	2.44			
Nivolumab	[redacted]	2.85	£10,978	0.40	£27,171
Dabrafenib	[redacted]	1.69	£19,602	-0.75	Dominated
Vemurafenib	[redacted]	1.70	£38,033	-0.74	Dominated

4.3.4 Including DTIC as a comparator in BRAF mutation-positive patients

The ERG observed that the company did not include DTIC as a comparator in BRAF mutation-positive patients, as discussed in Section 4.2.3 of this report. The ERG, therefore conducted a scenario analysis in which DTIC was included as one of the comparator arms, the results of which are presented in Table 32. In this scenario, nivolumab was the most cost-effective option with an ICER (nivolumab vs DTIC) of £21,201.

Table 32 - Including DTIC as a comparator arm in the BRAF mutation-positive analysis at list price

Treatment	Costs	QALYs	Incremental Costs (vs DTIC)	Incremental QALYs (vs DTIC)	ICER (£/QALY)
DTIC	[redacted]	1.10			
Ipilimumab	[redacted]	2.44	£53,793	1.35	Extendedly dominated
Nivolumab	[redacted]	4.27	£67,167	3.17	£21,201
Dabrafenib	[redacted]	1.69	£73,396	0.60	Dominated
Vemurafenib	[redacted]	1.70	£91,826	0.61	Dominated

4.3.5 Combination scenario with varying maximum treatment duration for nivolumab

The ERG conducted a combination scenario analysis whereby the following assumptions were simultaneously made to the cost-effectiveness model:

- Using a Weibull distribution for modelling TTP for nivolumab patients (the ERG considered this to be the best visual fit)
- Modelling PFS using the lognormal distribution for BRAF inhibitors (the ERG considered this to be the best visual fit)
- Using the data extrapolation method to model long-term survival for nivolumab
- Treatment duration ranging from two years to no maximum treatment duration.

The results of the combination scenarios (shown in Table 33, Table 34, Table 35,

Table 36 Table 37, and Table 38) show that nivolumab is dominated by ipilimumab in both BRAF mutation-negative and BRAF mutation-positive patients. The cost of nivolumab is almost double in the scenario with no maximum treatment duration compared to using maximum treatment duration of two years.

Table 33 - Combination scenario at list price (BRAF mutation-negative) 2 years treatment duration

Treatment	Costs	QALYs	Incremental Costs (vs DTIC)	Incremental QALYs (vs DTIC)	ICER (£/QALY)
DTIC	[redacted]	1.23			
Ipilimumab	[redacted]	2.64	£48,429	1.41	£34,261
Nivolumab	[redacted]	2.55	£69,725	1.32	Dominated

Table 34 - Combination scenario at list price (BRAF mutation-positive) 2 years treatment duration

Treatment	Costs	QALYs	Incremental Costs (vs ipilimumab)	Incremental QALYs (vs ipilimumab)	ICER (£/QALY)
Ipilimumab	[redacted]	2.44			
Dabrafenib	[redacted]	1.69	£4,860	-0.76	Dominated
Nivolumab	[redacted]	2.27	£5,267	-0.17	Dominated
Vemurafenib	[redacted]	1.70	£19,606	-0.75	Dominated

Table 35 - Combination scenario at list price (BRAF mutation-negative) 3 years treatment duration

Treatment	Costs	QALYs	Incremental Costs (vs DTIC)	Incremental QALYs (vs DTIC)	ICER (£/QALY)
DTIC	[redacted]	1.23			
Ipilimumab	[redacted]	2.64	£48,429	1.41	£34,261
Nivolumab	[redacted]	2.54	£84,257	1.31	Dominated

Table 36 - Combination scenario at list price (BRAF mutation-positive) 3 years treatment duration

Treatment	Costs	QALYs	Incremental Costs (vs ipilimumab)	Incremental QALYs (vs ipilimumab)	ICER (£/QALY)
Ipilimumab	[redacted]	2.44			£43,603
Dabrafenib	[redacted]	1.69	£4,860	-0.76	Dominated
Vemurafenib	[redacted]	1.70	£14,746	-0.75	Dominated
Nivolumab	[redacted]	2.26	£22,574	-0.18	Dominated

Table 37 - Combination scenario at list price (BRAF mutation-negative) maximum treatment duration

Treatment	Costs	QALYs	Incremental Costs (vs DTIC)	Incremental QALYs (vs DTIC)	ICER (£/QALY)
DTIC	[redacted]	1.23			
Ipilimumab	[redacted]	2.64	£48,429	1.41	£34,261
Nivolumab	[redacted]	2.51	£155,177	1.28	Dominated

Table 38 - Combination scenario at list price (BRAF mutation-positive) maximum years treatment duration

Treatment	Costs	QALYs	Incremental Costs (vs ipilimumab)	Incremental QALYs (vs ipilimumab)	ICER (£/QALY)
Ipilimumab	[redacted]	2.44			£43,603
Dabrafenib	[redacted]	1.69	£4,860	-0.76	Dominated
Vemurafenib	[redacted]	1.70	£14,746	-0.75	Dominated
Nivolumab	[redacted]	2.24	£83,858	-0.21	Dominated

4.4 Summary of uncertainties and issues

The CS reports that nivolumab is cost effective compared to its comparators at a cost effectiveness threshold of £30,000 per QALY and the base case results are robust to uncertainties of key model parameters and assumptions. However there are some uncertainties with regard to the modelling assumptions and data. The ERG notes that incorporating changes to the method used to estimate OS, the maximum treatment duration and TTP have significant impact on the model results. In the ERG combination scenario analysis, nivolumab is no longer cost effective and is dominated by ipilimumab. Furthermore, the ERG notes that a key trial,

CheckMate 067, has not been included in the company's analysis due to lack of available OS data.

5 End of life

The CS discusses the end of life criteria in Table 52 and states that advanced melanoma is associated with a short life expectancy, with median survival estimates of 6-10 months. Survival analyses of CheckMate 066 trial data indicate that nivolumab offers an extension to life of at least three months compared to palliative chemotherapy (DTIC). However, the survival benefit compared to ipilimumab is not yet fully established, pending follow-up OS data from CheckMate 067.[5] The CS reported that the expected number of new cases and relapsed cases of advanced melanoma in England in 2016 is 1,577. The CS therefore concluded that nivolumab is suitable for consideration as a life-extending treatment at the end of life.

The ERG also notes that in TA319[16] for ipilimumab for advanced melanoma, the Appraisal Committee was satisfied that ipilimumab met the criteria for being a life-extending, end of life treatment.

6 Innovation

The CS states that nivolumab should be considered innovative, representing a step-change in the management of advanced melanoma. The arguments in support of this include the stated significant clinical improvement associated with the drug, demonstrated through 45-50% of patients estimated to still be in remission two years after treatment initiation, based on extrapolation from the on-going Phase III RCTs. Furthermore, the CS reports that the Medicines and Healthcare products Regulatory Agency awarded nivolumab a Promising Innovative Medicine (PIM) designation for the treatment of advanced melanoma. Nivolumab was also approved to treat locally advanced or metastatic squamous non-small cell lung cancer (NSCLC) through the Early Access to Medicines Scheme. The criteria for drugs to be supported under this scheme include evidence that the product is likely to offer significant advantage over methods currently used in the UK.

7 DISCUSSION

7.1 Summary of clinical effectiveness issues

The clinical effectiveness evidence for nivolumab is based on three on-going phase III RCTs. The trials were conducted internationally, though a small proportion of UK patients were

included in two of them. The ERG considers the trials to be well-designed and unlikely to be at high risk of bias.

The currently available evidence shows that nivolumab is associated with a significant reduction in mortality compared to DTIC. However, the impact of nivolumab on overall survival compared to ipilimumab is not yet reported. Nivolumab also increased PFS compared to DTIC or ipilimumab. In terms of treatment response (ORR) there was significant benefit of nivolumab over comparator drugs in all three CheckMate trials. From the limited currently available nivolumab does not impair HRQoL. However, there is no current evidence that nivolumab leads to a consistent and sustained improvement in HRQoL. Nivolumab has a favourable AE profile with a lower incidence of high grade and serious AEs in compared to comparators, although nearly all trial participants experienced AEs (of any grade or category). Expert clinical advice to the ERG suggested that the benefits seen so far are very clinically significant.

A mixed treatment comparison of all comparators to inform economic modelling was not possible, necessitating an indirect comparison using selected RCTs from the company's systematic review of clinical effectiveness. Two separate evidence networks were created for BRAF mutation-positive and BRAF mutation-negative patients, respectively. A complex process was followed based on extraction of patient-level data from the trials ('pseudo patient-level data' from the BRAF inhibitor trials), using TTP, PrePS, and PPS outcomes to estimate PFS and OS (where available long-term data are currently unavailable). Covariate-adjusted parametric survival models, to adjust for differences between the trials, were created to inform transitions between states in the economic model. As summarised earlier in this report, the ERG considers that, in the circumstances, the approach taken was reasonable (subject to caveats for possible uncertainties, such as small sample sizes and numbers of events for some of the outcomes included), with some of the assumptions underpinning the indirect comparison having been accepted in previous NICE appraisals of treatments for advanced melanoma. However, there may be uncertainty around the assumption that there is no difference in treatment effect for nivolumab by BRAF mutation status. This is of significance because the cost-effectiveness estimates for BRAF-mutation positive patients are informed by the results of the CheckMate 066 trial which only included BRAF-mutation negative patients.

One of the biggest limitations was the omission of the pivotal CheckMate 067 trial from the indirect comparison evidence networks as this would have provided a direct comparison

between nivolumab and ipilimumab. The company clarified that it was not possible to have used data for the alternative outcomes from this trial (i.e. TTP, PrePS and PPS) as had been done for CheckMate 066 as this requires both PFS and OS events to be available. This appears to be a reasonable argument.

7.2 Summary of cost effectiveness issues

The CS includes evidence on the cost effectiveness of nivolumab compared with DTIC and ipilimumab for BRAF-mutation-negative patients and with dabrafenib, ipilimumab and vemurafenib for BRAF-mutation-positive patients with advance melanoma. The methods adopted for the economic evaluation are reasonable and are generally appropriate. The model structure and model parameter inputs are consistent with the clinical disease pathways and the available clinical trial evidence. However, the CS has not used direct evidence from the CheckMate 067 trial for nivolumab versus the ipilimumab and instead has conducted an indirect comparison. There is also some uncertainty regarding the maximum treatment duration for nivolumab.

The company performed a wide range of sensitivity analyses including one-way, probabilistic and scenario analyses to assess model uncertainty. Across most of the scenarios and sensitivity analyses, nivolumab was found to be cost effective in both BRAF mutation-positive and BRAF mutation-negative patients. The model results from the PSA suggest that in BRAF mutation-positive patients, the probabilities at list price were 100% at both £30,000 and £50,000 willingness-to-pay thresholds respectively and at estimated PAS prices, the probabilities were [redacted] respectively. In BRAF mutation-negative patients, the probabilities of nivolumab being cost-effective at list price were 87% and 99% at willingness-to-pay thresholds of £30,000 and £50,000 respectively and the probabilities for the PAS analyses were [redacted] respectively. The key model drivers in the one-way sensitivity analyses were: parameters that defined the fitted parametric curves for TTP, PPS, long-term OS; OS/PFS for vemurafenib; time on treatment; utility parameters; and administration cost.

The company has implemented two important assumptions: (i) that the long-term overall survival will be similar to seen with ipilimumab, i.e. an OS plateau, however this may not be the case and other distributions for long term OS may be more appropriate; and (ii) that the maximum treatment duration should be two years, although the marketing authorisation

specifies that treatment should continue as long as clinical benefit is observed or until treatment is no longer tolerated by the patient.

The ERG believes that the comparative efficacy of nivolumab with the comparator treatments in the NICE scope is uncertain due to a lack of head-to-head data from clinical trials. Furthermore changes to the method used to estimate OS, the maximum treatment duration and TTP have significant impact on the model results.

8 REFERENCES

(1) National Institute for Health and Care Excellence (NICE). Melanoma: Assessment and management. NICE, London . 2015. 7-10-2015.

(2) Bristol-Myers Squibb Company. Summary of Product Characteristics (SPC). OPDIVO 10 ml/ml concentratefor solution for infusion. 2015. 7-10-2015.

(3) National Institute for Health and CARE Excellence (NICE). Pembrolizumab for treating advanced melanoma after disease progression with ipilimumab. 2015. 7-10-2015.

(4) Robert C, Long GV, Brady B, Dutriaux C, Maio M, Mortier L et al. Nivolumab in previously untreated melanoma without BRAF mutation. *New England Journal of Medicine* 2015; 372(4):320-330.

(5) Larkin J, Chiarion-Sileni V, Gonzalez R, Grob JJ, Cowey CL, Lao CD et al. Combined Nivolumab and Ipilimumab or Monotherapy in Untreated Melanoma. *New England Journal of Medicine* 2015; 373(1):23-34.

(6) A cost effectiveness analysis of nivolumab compared to ipilimumab for the treatment of Braf Wild-Type Advanced Melanoma in Australia. ISPOR 18th Annual European Congress; 2015.

(7) Topalian SL, Hodi FS, Brahmer JR, Gettinger SN, Smith DC, McDermott DF et al. Safety, Activity, and Immune Correlates of Anti-PD-1 Antibody in Cancer. *New England Journal of Medicine* 2012; 366(26):2443-2454.

(8) Weber JS, Kudchadkar RR, Yu B, Gallenstein D, Horak CE, Inzunza HD et al. Safety, Efficacy, and Biomarkers of Nivolumab With Vaccine in Ipilimumab-Refractory or -Naive Melanoma. *Journal of Clinical Oncology* 2013; 31(34):4311-4U50.

(9) Weber JS, D'Angelo SP, Minor D, Hodi FS, Gutzmer R, Neyns B et al. Nivolumab versus chemotherapy in patients with advanced melanoma who progressed after anti-CTLA-4

treatment (CheckMate 037): a randomised, controlled, open-label, phase 3 trial. *Lancet Oncology* 2015; 16(4):375-384.

(10) Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer* 2009; 45(2):228-247.

(11) Winstanley JB, Young TE, Boyle FM, Bergenmar M, Bottomley A, Burmeister B et al. Cross-cultural development of a quality-of-life measure for patients with melanoma: phase 3 testing of an EORTC Melanoma Module. *Melanoma Res* 2015; 25(1):47-58.

(12) Chapman PB, Hauschild A, Robert C, Haanen JB, Ascierto P, Larkin J et al. Improved survival with vemurafenib in melanoma with BRAF V600E mutation. *N Engl J Med* 2011; 364(26):2507-2516.

(13) Hauschild A, Grob JJ, Demidov LV, Jouary T, Gutzmer R, Millward M et al. Dabrafenib in BRAF-mutated metastatic melanoma: a multicentre, open-label, phase 3 randomised controlled trial. *Lancet* 2012; 380(9839):358-365.

(14) Robert C, Thomas L, Bondarenko I, O'Day S, Weber J, Garbe C et al. Ipilimumab plus dacarbazine for previously untreated metastatic melanoma. *N Engl J Med* 2011; 364(26):2517-2526.

(15) Hodi FS, O'Day SJ, McDermott DF, Weber RW, Sosman JA, Haanen JB et al. Improved survival with ipilimumab in patients with metastatic melanoma. *N Engl J Med* 2010; 363(8):711-723.

(16) National Institute for Health and CARE Excellence (NICE). Ipilimumab for previously untreated advanced (unresectable or metastatic) melanoma. 2014.

(17) Larkin J, Lao CD, Urba WJ, McDermott DF, Horak C, Jiang J et al. Efficacy and Safety of Nivolumab in Patients With BRAF V600 Mutant and BRAF Wild-Type Advanced Melanoma: A Pooled Analysis of 4 Clinical Trials. *JAMA Oncol* 2015; 1(4):433-440.

(18) Overall survival in the management of pretreated patients with unresectable stage III/IV Melanoma: A systematic literature review and meta-analysis. The American Society of Clinical Oncology Annual Meeting. Chicago, IL, USA. 3-7 June 2011. Poster 8580; 2011.

(19) Meta-analysis of Kaplan-Meier overall survival curves from selected randomized controlled Phase II/III trials in advanced melanoma. Population Approach Group Europe Meeting. Athens, Greece. 7-10 June 2011; 2011.

(20) National Institute for Health and CARE Excellence (NICE). Ipilimumab for previously treated advanced (unresectable or metastatic) melanoma. 2012.

(21) Korn EL, Liu PY, Lee SJ, Chapman JA, Niedzwiecki D, Suman VJ et al. Meta-analysis of phase II cooperative group trials in metastatic stage IV melanoma to determine progression-free and overall survival benchmarks for future phase II trials. *J Clin Oncol* 2008; 26(4):527-534.

(22) Bucher HC, Guyatt GH, Griffith LE, Walter SD. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *J Clin Epidemiol* 1997; 50(6):683-691.

(23) Nikolakopoulou A, Chaimani A, Veroniki AA, Vasiliadis HS, Schmid CH, Salanti G. Characteristics of Networks of Interventions: A Description of a Database of 186 Published Networks. *PLoS One* 2014; 9(1):e86754.

(24) National Institute for Health and CARE Excellence (NICE). Dabrafenib for treating unresectable or metastatic BRAF V600 mutation-positive melanoma. 2014.

(25) Robert C, Karaszewska B, Schachter J, Rutkowski P, Mackiewicz A, Stroiakovski D et al. Improved overall survival in melanoma with combined dabrafenib and trametinib. *N Engl J Med* 2015; 372(1):30-39.

(26) Guyot P, Ades AE, Ouwens MJ, Welton NJ. Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan-Meier survival curves. *BMC Med Res Methodol* 2012; 12:9.

(27) Picot J, Copley V, Colquitt JL, Kalita N, Hartwell D, Bryant J. The INTRABEAM(R) Photon Radiotherapy System for the adjuvant treatment of early breast cancer: a systematic review and economic evaluation. *Health Technol Assess* 2015; 19(69):1-190.

(28) Effect of nivolumab (NIVO) on quality of life (QoL) in patients (pts) with treatment-naive advanced melnoma (MEL): results of a phase III study (CheckMate 066). American Society of Clinical Oncology Annual Meeting; Chicago, IL, USA: 2015.

(29) MIMS. Presciption drug database and drug prescribing guide. http://www.mims co uk/ [2015

(30) Department of Health. NHS reference costs 2013-2014. http://www.dh gov uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_122803 [14 A.D. [cited 2010 June 7];

(31) Curtis L. Unit Costs of Health and Social Care 2014. http://www.pssru.ac uk/pdf/uc/uc2010/uc2010 pdf [2014

(32) Oxford Outcomes. Melody study resource use data analysis: UK results. Data on file [2010

(33) Drummond MF, Jefferson TO. Guidelines for authors and peer reviewers of economic submissions to the BMJ. The BMJ Economic Evaluation Working Party. *BMJ* 1996; 313(7052):275-283.

(34) National Institute for Health and Care Excellence (NICE). Lapatinib or trastuzumab in combination with an aromatase inhibitor for the first-line treatment of metastatic hormone-receptor-positive breast cancer that overexpresses HER2. 2012.

(35) National Institute for Health and Care Excellence (NICE). Erlotinib for the first-line treatment of locally advanced or metastatic EGFR-TK mutation-positive non-small-cell lung cancer. 2012.

(36) National Institute for Health and Care Excellence (NICE). Bortezomib for induction therapy in multiple myeloma before high-dose chemotherapy and autologous stem cell transplantation. 2014.

(37) Balch CM, Buzaid AC, Soong SJ, Atkins MB, Cascinelli N, Coit DG et al. Final version of the American Joint Committee on Cancer staging system for cutaneous melanoma. *J Clin Oncol* 2001; 19(16):3635-3648.

(38) Schadendorf D, Hodi FS, Robert C, et al. Pooled analysis of long-term survival data from Phase II and Phase III trials of ipilimumab in unresectable or metastatic melanoma. *J Clin Oncol* 2015; 33(17):1889-1894.

(39) National Institute for Health and Care Excellence. Guide to the methods of technology appraisal 2013. *NICE, London* 2013.

(40) Centre for Reviews and Dissemination. Systematic reviews: CRD's guidance for undertaking reviews in health care. Third edition. 2009. York Publishing Services Ltd., CRD.

(41) Beusterien KM, Szabo SM, Kotapati S, Mukherjee J, Hoos A, Hersey P et al. Societal preference values for advanced melanoma health states in the United Kingdom and Australia. *Br J Cancer* 2009; 101(3):387-389.

(42) Department of Health. NHS reference costs 2013-2014. http://www.dh.gov uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_122803 [2014 [cited 2010 June 7];