in collaboration with:

ERASMUS UNIVERSITEIT ROTTERDAM
INSTITUTE OF HEALTH POLICY & MANAGEMENT

Maastricht University

Pembrolizumab for treating relapsed or refractory classical Hodgkin lymphoma

Produced by Kleijnen Systematic Reviews Ltd. in collaboration with Erasmus University Rotterdam (EUR) and Maastricht University

Authors
Debra Fayter, Systematic Reviewer, KSR Ltd, UK
Sabine Grimm, Health Economist, Maastricht UMC
Bram Ramaekers, Health Economist, Maastricht UMC
Svenja Petersohn, Health Economist, Maastricht UMC
Rob Riemsma, Reviews Manager, KSR Ltd, UK
Nigel Armstrong, Health Economist, KSR Ltd, UK
Xavier Pouwels, Health Economist, Maastricht UMC
Willem Witlox, Health Economist, Maastricht UMC
Caro Noake, Information Specialist, KSR Ltd, UK
Gill Worthy, Statistician, KSR Ltd, UK
Manuela Joore, Health Economist, Maastricht UMC
Jos Kleijnen, Director, KSR Ltd, Professor of Systematic Reviews in Health Care, Maastricht University

Correspondence to
Rob Riemsma, Kleijnen Systematic Reviews
Unit 6, Escrick Business Park
Riccall Road, Escrick
York, UK
YO19 6FD

Date completed 14/11/2017

Source of funding: This report was commissioned by the NIHR HTA Programme as project number STA 16/119/02.

Declared competing interests of the authors
None.

Acknowledgements
None

[redacted]

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:
Fayter D, Grimm S, Ramaekers B, Petersohn S, Riemsma R, Armstrong N, Witlox W, Pouwels X, Noake C, Worthy G, Joore M, Kleijnen J. Pembrolizumab for treating relapsed or refractory classical Hodgkin lymphoma; A Single Technology Assessment. York: Kleijnen Systematic Reviews Ltd, 2017.

Contributions of authors
Debra Fayter acted as project lead and systematic reviewer on this assessment, critiqued the clinical effectiveness methods and evidence and contributed to the writing of the report. Sabine Grimm acted as health economic project lead, critiqued the company's economic evaluation and contributed to the writing of the report. Bram Ramaekers, Svenja Petersohn, Xavier Pouwels, Willem Witlox and Nigel Armstrong acted as health economists on this assessment, critiqued the company's economic evaluation and contributed to the writing of the report. Rob Riemsma acted as a systematic reviewer, critiqued the clinical effectiveness methods and evidence and contributed to the writing of the report. Gill Worthy acted as statistician, critiqued the analyses in the company's submission and contributed to the writing of the report. Caro Noake critiqued the search methods in the submission and contributed to the writing of the report. Manuela Joore critiqued the manufacturer's economic evaluation, contributed to the writing of the report and provided general guidance. Jos Kleijnen critiqued the company's definition of the decision problem and their description of the underlying health problem and current service provision, contributed to the writing of the report and supervised the project.

Abbreviations

ABVD regimen	Doxorubicin, bleomycin, vinblastine and dacarbazine
AE	Adverse Events
AEOSI	Adverse events of special interest
AIC	Akaike Information Criterion
AlloSCT	Allogeneic Stem Cell Transplant
ASaT	All Subjects as Treated
ASCO	American Society of Clinical Oncology
ASHAP	Doxorubicin, methylprednisolone, cytarabine, cisplatin
AUC	Area under the curve
AutoSCT	Autologous Stem Cell Transplant
BCSH	British Committee for Standards in Haematology Guidelines
BEACOPP regimen	Bleomycin, etoposide, doxorubicin, cyclophosphamide, vincristine, procarbazine and prednisone
BIC	Bayesian information criterion
BICR	Blinded independent central radiologists'
BNF	British National Formulary
BOR	Best overall response
BSA	Body surface area
BSC	Best supportive care
BTD	Breakthrough Therapy Designation
BV	Brentuximab Vedotin
C	Cirrhotic
CADTH	Canadian Agency for Drugs and Technologies in Health
CAA	Commercial access agreement
CDF	Cancer Drugs Fund
CE	Cost Effectiveness
CEA	Cost effectiveness Analysis
CEAC	Cost effectiveness Acceptability Curve
cHL	Classical Hodgkin Lymphoma
CHMP	Committee for Medicinal Products for Human Use
CHOP	cyclophosphamide, doxorubicin, prednisolone, vincristine
CI	Confidence Interval
CMU	Commercial medicines unit
CPS	Combined positive score
CR	Complete response
CRD	Centre for Reviews and Dissemination
CrI	Credible interval
CSR	Clinical study report
CT	Computer Tomography
CTCAE	Common Terminology Criteria for Adverse Events
DAA	Direct-acting antivirals
DAE	Discontinuation due to adverse events
DHAOx	Dexamethasone, cytarabine, oxaliplatin
DHAP	Dexamethasone, cytarabine, cisplatin
DoH	Department of Health
DOR	Duration Of Response
DSU	Decision Support Unit
ECOG	Eastern Cooperative Oncology Group
EGFR	Epidermal Growth Factor Receptor
EMA	European Medicines Agency
EPAR	European public assessment report
EORTC-QLQC30	European Organisation for Research and Treatment Cancer Quality of Life Questionnaire

EQ-5D	European Quality of Life-5 Dimensions
ERG	Evidence Review Group
ESHAP	Etoposide, methylprednisolone, cytarabine, cisplatin
ESMO	European Society for Medical Oncology
EUR	Erasmus University Rotterdam
FAS	Full analysis set
FDA	Food and Drug Administration
GDP	Gemcitabine, dexamethasone, cisplatin
GEM-P	Gemcitabine, cisplatin, methylprednisolone
GVD	Gemcitabine, vinorelbine, liposomal doxorubicin
GVHD	Graft Versus Host Disease
HL	Hodgkin Lymphoma
HR	Hazard ratio
HRG	Healthcare Resource Group
HRQoL	Health-related Quality of Life
HTA	Health Technology Assessment
IC	Indirect Comparison
ICE	Ifosfamide, carboplatin, etoposide
ICER	Incremental Cost-effectiveness Ratio
ICTRP	International Clinical Trials Registry Platform
IRG	Independent Review Group
ISPOR	International Society for Pharmacoecomics and Outcomes Research
ITT	Intention to Treat
IV	Intravenous
IVE	Ifosfamide, epirubicin, etoposide
IVOx	Ifosfamide, etoposide, oxaliplatin
IWG	International Working Group
KM	Kaplan-Meier
KSR	Kleijnen Systematic Reviews
LY	Life Year
mAB	Monoclonal antibody
MAIC	Matched Adjusted Indirect treatment comparison
MeSH	Medical Subject Headings
MHRA	Medicines and Healthcare Products Regulatory Agency
MINE	Mitoxantrone, ifosfamide, vinorelbine, etoposide
MK-3475	Pembrolizumab - Keytruda®
MS	Manufacturer's Submission
MSD	Merck Sharp and Dohme
MTC	Mixed Treatment Comparison
NA	Not applicable
NHL	Non Hodgkin Lymphoma
NHS	National Health Services
NICE	National Institute for Health and Care Excellence
NIHR	National Institute for Health Research
NMA	Network meta-analysis
NR	Not Reported
ORR	Objective Response Rate
OS	Overall survival
PD	Progressive Disease
PD-1	Programmed death 1 protein
PD-L1	Programmed death ligand 1
PET	Positron Emission Tomography
PFR	Progression-free rate
PFS	Progression free survival
PI	Principal Investigator

PIM	Promising Innovative Medicines
PK	Pharmacokinetics
PMitCEBO	Bleomycin, cyclophosphamide, etoposide, mitoxantrone, prednisolone, vincristine
PR	Partial response
PRESS	Peer Review of Electronic Search Strategies
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRO	Patient-reported outcome
PSA	Probabilistic Sensitivity Analyses
PSS	Personal Social Services
PSSRU	Personal and Personal and Social Services Research Unit
Q3W	Every 3 weeks
QALY(s)	Quality-adjusted Life Year(s)
QoL	Quality of life
RCT	Randomised Controlled Trial
RR	Response Rate; Relative Risk; Risk Ratio
RRcHL	Relapsed or refractory classical Hodgkin Lymphoma
RSC	Reed-Sternberg cells
RVIG	Gemcitabine, ifosfamide, mesna, prednisolone, rituximab, vinorelbine
SAE	Serious Adverse Events
SCT	Stem cell transplant
SD	Stable Disease; Standard deviation
SG	Standard Gamble
SIGN	Scottish Intercollegiate Guidelines Network
SLR	Systematic literature review
SMC	Scottish Medicines Consortium
SmPC	Summary of product characteristics
SOC	Standard of Care
STA	Single Technology Appraisal
TA	Technology Appraisal
ToT	Time on Treatment
TTO	Time trade off
UK	United Kingdom of Great Britain and Northern Ireland
UMC	University Medical Centre
VAS	Visual Analogue Scale
VAT	Value-Added Tax

Table of Contents

Table of Tables

Table of Figures

1. SUMMARY

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

The population of this appraisal is in line with the NICE scope. The main trial in the company submission (CS) (KEYNOTE-087) covers both cohorts of interest (cohort 1: people with relapsed or refractory classical Hodgkin Lymphoma (cHL) who have received autologous stem cell transplant (autoSCT) and brentuximab vedotin (BV) and, cohort 2: those who have received BV when autoSCT is not a treatment option). However only 14 patients in the trial were from the UK. None of the patients in the comparator study (Cheah et al. 2016) were from the UK. The comparator study in this appraisal was also used in a previous appraisal (TA462). NICE concluded in TA462 that "the comparator data may not fully represent UK clinical practice".

The intervention (pembrolizumab) is in line with the scope. Regulatory approval by the EMA for the indication considered within this submission was granted on the 2nd May 2017.

The description of the comparators in the NICE scope is as follows:

- Single or combination chemotherapy including drugs such as gemcitabine, vinblastine and cisplatin
- Best supportive care.

The company uses one retrospective USA database study as a comparator. In this study patients received the following types of therapy: investigational agent(s), gemcitabine, bendamustine, any other alkylator, BV retreatment, platinum based treatment, autoSCT or allogeneic SCT (alloSCT), or other treatment. The company has not provided separate data for comparators; instead a combined data set has been provided for multiple comparators, some of which are within the scope and others not. This combined data set was used as a comparator for both populations, cohort 1 and cohort 2.

The company's submission matches the NICE scope on outcome measures. The primary outcome in the KEYNOTE-087 trial is overall response rate (ORR). Although progression-free survival and overall survival are investigated, as per the NICE scope, the data for these outcomes are not fully mature.

1.2 Summary of clinical effectiveness evidence submitted by the company

The company did not identify any randomised controlled trials of pembrolizumab and its comparators in patients with classical Hodgkin Lymphoma who have either received autoSCT and BV or BV alone due to autoSCT being unsuitable. One ongoing, single arm study of the efficacy and safety of pembrolizumab was identified (KEYNOTE-087) and this formed the basis of the submission. KEYNOTE-087 includes 150 patients (14 UK patients) relevant to this appraisal. It covers both cohorts of interest (cohort 1: people with relapsed or refractory cHL who have received autologous stem cell transplant and brentuximab vedotin and cohort 2: patients who have received brentuximab vedotin when autologous stem cell transplant is not a treatment option). The company presented data based on a median follow up of 15.9 months. The median time on treatment was [redacted] days for cohort 1 and [redacted] days for cohort 2.

The primary outcome of KEYNOTE-087 was overall response rate (ORR) as assessed by independent committee. ORR was 75.4% in cohort 1 and 66.7% in cohort 2 over the course of the trial. Median progression free survival (PFS) in cohort 1 was 16.7 months (11.2 to NR). In cohort 2 it was 11.1 months (7.6 to 13.7). Median overall survival (OS) was [redacted]. At 12 months survival was [redacted] in cohort 1 and [redacted] in cohort 2. In cohort 1 [redacted] of patients had one or more adverse events. In cohort 2 [redacted] of patients had one or more adverse events. The company noted that

most AEs were low grade ([redacted] and [redacted] grades 3 to 5 in cohort 1 and 2 respectively). In cohort 1 [redacted] of AEs were classed as serious and in Cohort 2 [redacted]. The most common adverse events were pyrexia, cough, fatigue, diarrhoea and vomiting. The company conducted post-hoc analyses of response at 12 weeks to use in the comparison of clinical and cost effectiveness. Overall response rates were lower at 12 weeks than over the whole course of the trial ([redacted]

As KEYNOTE-087 did not have a comparator group the company identified a comparative observational study from the literature (Cheah 2016 et al). This is a retrospective USA database study in which patients received the following types of therapy: investigational agent(s), gemcitabine, bendamustine, any other alkylator, BV retreatment, platinum based treatment, autoSCT or alloSCT, or other treatment. The company has not provided separate data for comparators; instead a combined data set has been provided for multiple comparators.

The company performed two types of analyses: a naïve indirect comparison between KEYNOTE-087 and Cheah and a matched adjusted indirect treatment comparison (MAIC) of the two studies. With the exception of one of the naïve comparisons, all results significantly favoured pembrolizumab over SoC for ORR and PFS.

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

The CS and response to clarification provided sufficient details for the ERG to appraise the searches for eligible trials and to critique the clinical effectiveness of pembrolizumab. Searches were carried out in accordance with the NICE guide to the methods of technology appraisal Sections 5.2.2 and 5.2.4 using a good range of databases. Additional searches of conference proceedings, trials database and the NICE website were reported.

The clinical effectiveness of pembrolizumab submitted in this appraisal is based on the KEYNOTE-087 study. The most important methodological aspect to note is that although this trial was well conducted, it represents a low level of evidence. It is a phase II, single arm, non-comparative study which by its design has serious limitations. We cannot know whether the outcomes observed are a true reflection of the intervention as the role of natural history and baseline characteristics is not taken into account. As treatment is known to participants, clinicians and assessors this can lead to bias in the delivery of the intervention and the reporting of outcomes. Other limitations in applying the results of the trial to UK practice include the fact that median progression free survival data were immature and median overall survival [redacted]. The trial has only 150 relevant participants so the evidence base for this appraisal is small. Patients over 65 are not well represented. Furthermore, a small number of patients were from the UK (14) so the trial may not totally reflect the UK population and setting. It is recognised, however, that the population matching the scope of this appraisal from which to draw participants is in itself small. In clinical practice, for those who are suitable, pembrolizumab represents a bridge to alloSCT, a potentially curative treatment. However, the company submission stated that 'KEYNOTE-087 was not designed as a 'bridging' study, therefore the uptake of alloSCT was very low overall across cohorts 1 and 2.' The company further stated that 'the use of stem cell transplant would have been at the discretion of the treating physician on a per patient basis.'

The main comparative study is a US observational study with a range of different treatments both within and outside the NICE scope. In the previous appraisal of nivolumab (TA462), the committee considered that "the Cheah study was the best available evidence for standard of care and considered it appropriate for its decision-making, but overall the clinical effectiveness of nivolumab compared with standard of

care was highly uncertain because the comparator data may not fully represent UK clinical practice." However, the ERG is not aware of a more appropriate source of data for the comparator population for this appraisal.

The ERG identified problems with compatibility of the two studies in the CS regarding baseline characteristics and methods of outcomes assessment. In the MAIC the company adjusted for potential confounding variables so that the KEYNOTE-087 study more closely resembled the Cheah study. According to DSU report 18, unanchored indirect comparisons (i.e. those based on single-arm studies) are susceptible to large amounts of systematic error unless all prognostic variables and effect modifiers are accounted for. However, in the current MAIC the company was dependent on the variables reported in Cheah et al. (2016) and these are unlikely to be all relevant prognostic variables and effect modifiers. Therefore, the results are likely to contain systematic error but it is not possible to estimate the size of the potential error. Both the naïve IC and MAIC have major limitations for decision making.

1.4 Summary of cost effectiveness submitted evidence by the company

The company developed a de novo cohort state transition model with health states based on response, uptake of alloSCT, and survival. The model structure consists of a short term component (first 12 weeks), a subsequent decision tree element (at 12 weeks) to determine the proportion of patients transiting to alloSCT (conditional upon response at 12 weeks) and a long-term component (after the first 12 weeks) separately for patients who had alloSCT and patients who did not have alloSCT at 12 weeks. At 12 weeks, patients were allocated to alloSCT based on their response status and probabilities of alloSCT uptake were applied conditional on patients' response status. Any alloSCTs were assumed to happen at this 12-week time point, without any lag. Justifying their approach, the company believed that alloSCT data from KEYNOTE-087 were not reflective of UK clinical practice and that they did not have Kaplan-Meier data for time-to-alloSCT from Cheah et al.

In line with the marketing authorisation and the final scope issued by NICE, two distinct populations were considered in the cost effectiveness model: patients with RRcHL who have failed autoSCT and BV (cohort 1) and patients with RRcHL who are autoSCT ineligible and have failed BV (cohort 2).

Pembrolizumab monotherapy is implemented as per its EMA Summary of Product Characteristics (SmPC) posology and method of administration for RRcHL (i.e. administered intravenously at a fixed dose of 200 mg over 30 minutes every three weeks [Q3W]). The company assumed that in the model pembrolizumab monotherapy will be provided for a maximum of 24 months (35 cycles).

The company only considered "standard of care" (SoC) as comparator in its base-case. SoC as considered by the company consists of the following regimens: chemotherapy, bendamustine or investigational agents. The distribution of patients among these regimens was based on the distribution observed in Cheah et al (2016). The company also presented a scenario analysis, in which best supportive care (BSC) was added as a comparator. The company justified this deviation from the scope (i.e. not including BSC in its base-case) with their belief that BSC use would be minimal as eligible patients are likely to receive therapy whenever feasible.

The model adopts the perspective of the NHS and Personal and Social Services (PSS) in England and Wales. The cycle length was one week to account for the length of treatment cycles. A half-cycle correction was applied. A time horizon of 40 years was adopted to capture all relevant costs and outcomes. All costs and utilities were discounted at a rate of 3.5% per year.

Treatment effectiveness for pembrolizumab was primarily based on the KEYNOTE-087 study. The primary data source for the SoC comparator was the Cheah et al (2016) study. The naïve indirect

treatment comparison was used to inform relative overall survival (OS), progression-free survival (PFS) and response rates at week 12. The MAIC was only used in scenario analysis. Both KEYNOTE-087 cohorts were compared with the Cheah et al (2016) study cohort.

Due to the company's model structure, treatment effectiveness and time to treatment discontinuation (TTD) were estimated for the pre-12 week period and for the post-12 week period separately. Parametric models were fitted to the entire study data from KEYNOTE-087 to estimate OS and PFS for patients receiving pembrolizumab in the pre-12 week period. To inform the decision tree element at week 12, response rates from KEYNOTE-087 were used, as well as two clinician surveys to inform estimates of alloSCT uptake conditional on response status. For the post-12 week period, treatment effectiveness depended on whether patients received alloSCT or not. Mortality post-alloSCT was based on Lafferty et al (2017) and post-progression mortality for patients who did not receive alloSCT was based on Cheah et al (2016). The company justified the use of different data sources by stating that survival data from KEYNOTE-087 were immature.

TTD for patients treated with pembrolizumab for the pre-12 week period was assumed to be equivalent to PFS. TTD for the post-12 week period was estimated directly from KEYNOTE-087. Furthermore, TTD for SoC was assumed equivalent to PFS in Cheah et al. TTD for pembrolizumab was capped at 24 months.

Health-related quality-of-life (HRQoL) was measured in KEYNOTE-087 at different time points, but only responses from week 12 were used to obtain health state utility values, ignoring observations at other time points. The company calculated utility values stratified by response and response rates at 12 weeks to obtain progression-free health state utilities, and used response rates from Lafferty et al to calculate the post-alloSCT utility. The company did not use the progressed disease utility score from KEYNOTE-087 and instead opted to use a utility decrement from Swinburn et al (2015).

The electronic market information tool (eMit) was used to acquire drug acquisition costs of pembrolizumab and components of SoC. When these were unavailable, costs from the British National Formulary were used. Administration costs were obtained from the NHS reference costs. The list price of 200 mg pembrolizumab was £5,260. Through a Commercial Access Agreement (CAA), [redacted]. The cost for SoC was assumed to consist of acquisition and administration costs for the different chemotherapy regimens (equal use assumed), and bendamustine. Health state costs consisted of monitoring costs and outpatient attendance. For the post-alloSCT health state, a one-off cost was applied.

In the deterministic base-case analysis, total QALYs and LYs gained as well as total costs (with the CAA) were larger in the pembrolizumab treatment arm compared to UK SoC in both cohorts. Incremental costs mainly stemmed from differences in acquisition costs and alloSCT costs between pembrolizumab and SoC. Pembrolizumab treatment resulted in deterministic incremental cost effectiveness ratios (ICERs) of £43,511 and £48,571 per QALY gained for cohort 1 and cohort 2 respectively, as per the company's corrected base-case.

The company performed probabilistic sensitivity analysis (PSA) and deterministic sensitivity analyses (DSA). The PSA with 1,000 iterations resulted in ICERs of £43,653 and £50,894 per QALY gained for cohorts 1 and 2 respectively for pembrolizumab versus SoC. The explored scenarios resulted in significant changes to the ICERs in both cohorts.

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

In the absence of cost effectiveness studies performed on the population and intervention of interest from the literature, the ERG agreed that a de novo approach to modelling cost effectiveness of pembrolizumab was necessary. However, it was unclear why the company did not provide a complete overview of the publications included and excluded from their cost effectiveness, cost and resource and utility and HRQoL systematic literature reviews (SLRs). The company prioritised aligning their sources with TA462 over using the results of their SLRs.

No justification was provided for the model structure only allowing patients to have alloSCT at 12 weeks after starting treatment, thereby ignoring responses that can occur at later time points (as acknowledged by the company). The alloSCT at 12 weeks assumption furthermore neglects the time required to identify a donor and schedule the procedure. This entails that alloSCT in the present model is performed earlier than would be expected in clinical practice. Hence, the post-alloSCT benefits are applied earlier, which favours pembrolizumab. The company failed to include a post-alloSCT progressed disease health state in their model, not in line with evidence from Lafferty et al, thereby also favouring pembrolizumab.

The populations described by the company are consistent with the final scope issued by NICE for this appraisal. For KEYNOTE-087, the company was able to distinguish between patients who did and did not receive autoSCT (i.e. cohort 1 and 2 respectively). The company did not have access to the individual patient level data in Cheah et al and hence used the mixed population for comparisons with both cohorts. This likely resulted in comparisons of pembrolizumab with SoC that may be favourable and non-favourable for pembrolizumab in cohorts 1 and 2 respectively.

BSC was not incorporated in the CS base-case (inconsistent with the scope), but only presented in a scenario analysis. Moreover, nivolumab was recently recommended by NICE in part of this population (cohort 1) and may become a relevant comparator in the future.

The assumption that pembrolizumab monotherapy will be stopped after 24 months is inconsistent with the SmPC but in line with the KEYNOTE-087 protocol. It is unclear whether pembrolizumab, in UK clinical practice, would also be provided for a maximum of 24 months. Removing this cap resulted in substantially increased ICERs for both cohorts, showing that the company's base-case might under-estimate the cost incurred with the use of pembrolizumab if a 24-months stopping rule is not enforced in clinical practice.

The ERG considered the adopted perspective and discounting to be appropriate for this appraisal.

Treatment and relative treatment effectiveness used in the model relied on the use of evidence from single-arm studies and a naïve indirect comparison. There is therefore substantial uncertainty about relative treatment effectiveness. The use of the naïve indirect comparison instead of the MAIC favoured SoC.

The combining of survey results to inform alloSCT uptake conditional on response status was viewed as inappropriate considering that the company acknowledged that it was possible for both surveys to include the same clinical experts. The company omitted the result from its survey that patients with progressed disease could still be eligible for alloSCT. Both assumptions favoured pembrolizumab.

Post-12 week mortality data from KEYNOTE-087 was deemed immature by the company and the ERG agreed with this assessment. [redacted] and the ERG considers that these may be informative for the present model. Furthermore, the ERG was

concerned about the use of Lafferty et al, given its small sample size and the questionable generalisability to UK clinical practice. The company's method used for extrapolating OS post-alloSCT was deemed by the ERG to over-estimate OS, which favoured pembrolizumab. There was also significant uncertainty around extrapolating PFS post-12 weeks, which translated into significant increases in the ICERs when alternative parametric survival models were chosen in both cohorts.

The mixed effects model utilities provided in response to the clarification letter, were deemed by the ERG to make better use of the KEYNOTE-087 data. The ERG preferred estimating the progressed disease (PD) utility from KEYNOTE-087, rather than Swinburn et al. The ERG considered the proportion of responders used for calculating utility values as inconsistent.

The ERG was concerned about the assumption that all chemotherapy agents contributed equally to the mix of SoC in calculating costs. This likely favoured pembrolizumab. Resource use and costs associated with alloSCT were likely under-estimated in the model, also favouring pembrolizumab.

Cost effectiveness results were not presented for BSC in the base-case. The number of iterations (1,000) in the PSA was likely too small to achieve stable results.

The ERG also had concerns about model validation, mostly relating to the lack of cross-validation with TA462 and the irreproducibility of model estimates used for external validation.

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

1.6.1 Strengths

Overall, the CS reported searches were well presented and easily reproducible. Searches were carried out on a good range of databases. The clinical effectiveness strategies utilised a recognised study design filter. Supplementary searches of conference proceedings and the NICE website were undertaken by the company, along with a manual search of the WHO ICTRP trial database in order to identify additional on-going trials. The clinical evidence is based on a well conducted, multicentre single-arm trial reflecting both cohorts of patients relevant to the decision problem. Outcomes assessed reflect the scope.

Overall, the model is well built and transparent. The company reflected that pembrolizumab can be considered as a bridging treatment to alloSCT by incorporating alloSCT in the economic model. The company provided alternative data (for example derived from the MAIC) and alternative survival functions to enable exploratory analyses in the model.

1.6.2 Weaknesses and areas of uncertainty

The ERG had some concerns about the language bias of restricting clinical effectiveness searches to English language only as this is not in line with current best practice. However, the main weakness of this appraisal is the lack of relevant randomised controlled trials (RCTs). Outcomes relating to pembrolizumab are based on a single arm trial. Comparisons with the comparators in the scope are problematic due to the availability of only one US study with a mix of different treatments. The naïve and matched adjusted comparisons conducted by the company have a number of limitations and represent a much weaker level of evidence than a RCT. Additionally progression-free survival and overall survival data are not fully mature. KEYNOTE-087 is an ongoing trial so more information will be available in future regarding uncertainties in progression-free and overall survival.

The model structure did not appropriately reflect the timing of the alloSCT decision and the timing of the actual alloSCT procedure. The model therefore under-estimates the time to alloSCT and assumes

that any benefits will be obtained sooner than is likely to occur in clinical practice. Furthermore, the company's model assumed that no patients would progress after receiving alloSCT. These assumptions favour pembrolizumab.

The company informed alloSCT uptake conditional on response status at 12 weeks after treatment start through a UK clinician survey and then combined these survey results with the previously performed BMS survey results (from TA462). The appropriateness of combining both surveys is questionable. The appropriate approach for incorporating alloSCT in the model would have been to use time to alloSCT data directly from the main source of evidence. There remains major uncertainty about the alloSCT uptake estimates. Furthermore, the elicited alloSCT uptake (from the MSD survey) for patients with progressed disease was ignored. Both, the combining of both surveys and ignoring alloSCT uptake in progressed disease patients, were shown in scenario analysis to be major drivers of cost effectiveness.

A major limitation was the use of single-arm evidence to inform treatment effectiveness. There was uncertainty whether the MAIC or the naïve indirect comparison should be used. The company provided both and the ERG, like the company, used the naïve indirect comparison in the base-case and the MAIC in scenario analysis. Furthermore, the ERG viewed the immaturity of the OS data from KEYNOTE-087 as a major limitation as this necessitated the use of post-alloSCT OS and utility estimates from alternative data sources, one of which was based on 13 patients only. The methods used to extrapolate from this data source were also questionable. [redacted], and the ERG considers that these may be informative for the present analysis.

It is of note that the population used for the comparator was a mixed population of cohorts 1 and 2, that is, it included patients who did and did not receive autoSCT, derived from Cheah et al. The Cheah et al. population is more comparable with KEYNOTE-087 cohort 1 than with cohort 2 in terms of patient characteristics. The use of this mixed comparator population likely resulted in comparisons of pembrolizumab with SoC that may be favourable and non-favourable for pembrolizumab in cohorts 1 and 2 respectively, but this could not be formally explored in scenario analysis.

Of further note, the economic model, and the evidence from KEYNOTE-087, rely on the assumption that treatment with pembrolizumab is capped at 24 months, which is inconsistent with its SmPC. It is unclear whether in UK clinical practice pembrolizumab would also be provided for a maximum of 24 months. This assumption favoured pembrolizumab.

Model extrapolations lack face and external validity. For example, the company claims that End of Life criteria can be considered fulfilled, however, their model predicts life year gains of 53 months on standard of care.

1.7 Summary of exploratory and sensitivity analyses undertaken by the ERG

A number of issues were identified by the ERG. The ERG was able to adjust/correct some of these in its base-case. This resulted in ICERs (probabilistic) of pembrolizumab (with confidential access agreement (CAA)) versus SoC of £64,186 and £78,696 per QALY gained for cohorts 1 and 2 respectively.

Additional sensitivity analyses were performed to examine the potential impact of alternative assumptions on the cost effectiveness estimates. The scenarios with the largest impact were alternative assumptions for extrapolating post-alloSCT, an alternative survival model for extrapolating post-12-week PFS in cohort 2, the use of the MAIC instead of the naïve comparison and removing the cap of 24 months on TTD (Table 1.1).

Table 1.1. ERG base-case and exploratory analyses

	Technologies	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	Pembrolizumab ICER (£/QALY)
Company corrected base-case cohort 1	Pembrolizumab	£107,459	4.497			
	SoC	£52,017	3.223	£55,442	1.274	£43,511
ERG base-case cohort 1	Pembrolizumab	£107,998	4.460			
	SoC	£50,913	3.535	£57,085	0.925	£61,705
Use of MAIC (2) cohort 1	Pembrolizumab	£107,998	4.460			
	SoC	£47,997	3.359	£60,001	1.102	£54,466
No 24-months cap on TTD (3) cohort 1	Pembrolizumab	£123,990	4.460			
	SoC	£50,913	3.535	£73,077	0.925	£78,992
Alternative OS post-alloSCT assumption (5)	Pembrolizumab	£107,030	3.558			
	SoC	£50,157	2.830	£56,873	0.727	£78,204
Company corrected base-case cohort 2	Pembrolizumab	£93,732	4.072			
	SoC	£51,424	3.200	£42,308	0.871	£48,571
ERG base-case cohort 2	Pembrolizumab	£93,095	4.118			
	SoC	£50,609	3.541	£42,486	0.577	£73,594
Alternative distributions (1.b) cohort 2	Pembrolizumab	£92,556	3.995			
	SoC	£50,550	3.529	£42,007	0.466	£90,152
Use of MAIC (2) cohort 2	Pembrolizumab	£93,095	4.118			
	SoC	£45,924	3.337	£47,171	0.781	£60,372
No 24-months cap on TTD (3) cohort 2	Pembrolizumab	£96,380	4.118			
	SoC	£50,609	3.541	£45,771	0.577	£79,284
Alternative OS post-alloSCT assumption (5)	Pembrolizumab	£92,204	3.287			
	SoC	£49,863	2.844	£42,341	0.442	£95,712

ERG = Evidence Review Group; ICER = incremental cost effectiveness ratio; QALY = quality-adjusted life year

2. BACKGROUND

In this report the ERG provides a review of the evidence submitted by Merck Sharp & Dohme (MSD) in support of pembrolizumab, trade name KEYTRUDA®, for the treatment of patients with relapsed or refractory Classical Hodgkin lymphoma (cHL). In this section we outline and critique the company's description of the underlying health problem and the overview of current service provision. The information is taken from Chapter 3 of the company submission (CS)[1] with sections referenced as appropriate.

2.1 Critique of company's description of underlying health problem

The underlying health problem of this appraisal is Classical Hodgkin Lymphoma which the company describes as 'a rare, localised or disseminated, malignant proliferation of cells of the lymphoreticular system, occurring mostly in lymph node tissues, spleen, liver and bone marrow.'[1]

The CS clarifies that Classical Hodgkin Lymphoma is the predominant subgroup of Hodgkin Lymphoma and accounts for 95% of cases of the disease. The presence of Reed-Sternberg cells in Hodgkin Lymphoma is highlighted.

There are four subtypes of Classical Hodgkin Lymphoma: nodular sclerosing (60%) which is usually identified early due to swelling of the lymph nodes in the neck; lymphocyte rich (20%), mixed cellularity (15%) and lymphocyte depleted (very rare).[2] Patients may present with bulky disease. This is defined as a lymph node that is 10cm or more or a lymphoma in the centre of the chest (mediastinum) which is at least one third of the width of the chest.[2]

The company highlights the symptomatic burden of cHL and that patients with B symptoms (presence of fever, weight loss and drenching night sweats) are associated with worse outcomes.

The CS cites Cancer Research UK data that states that in 2014 there were 2,106 new cases of Hodgkin Lymphoma in the UK. The CS also states that according to Cancer Research UK data incidence rates may increase by 5% in the UK population overall between 2014 and 2035.[1]

The company highlights that incidence of Hodgkin Lymphoma peaks in young adults (20 to 24 years of age) and older males and females (75 to 79 years of age) with approximately half of diagnoses reported in people aged 45 and over.

The company describes the survival rates for HL as promising with rates of 91.4% at one year, 85.0% at five years and 80.4% at 10 years. However, they caution that the relapsed/refractory population under consideration for this appraisal are likely to have a poorer prognosis compared with patients who respond to therapy. The company mention a retrospective trial of 81 patients showing a five year survival of less than 20%.[3]

The CS refers to the burden of costs affecting patients, caregivers and society. It is noted that there is a relatively high proportion of patients with Hodgkin Lymphoma who are of working age.

ERG comment:

- The company provides a good overview of the underlying health problem. The ERG checked the references provided to support the statements in the company submission. In general, these were found to be appropriate.
- The population in this appraisal is specifically people with relapsed or refractory cHL who have received autologous stem cell transplant (autoSCT) and brentuximab (BV) or BV when autoSCT is not a treatment option.

2.2 Critique of company's overview of current service provision

The company correctly reports that there is no NICE guideline on relapsed/refractory CHL.

For first line therapy chemotherapy alone or chemotherapy combined with radiotherapy is used in practice. Between 15 and 30% of patients with HL do not achieve remission with these treatments.[4] The CS outlines that those patients who do not achieve remission may be offered chemotherapy and/or radiotherapy to enable autoSCT. AutoSCT is potentially curative and effective in about 50% of people.[4] However autoSCT may not be an option for some patients if their disease does not respond adequately to treatment or the patient's age or comorbidities prevent offering it as an option.

The CS highlights the recent approval of brentuximab vedotin (BV) for patients with relapsed or refractory disease after autoSCT or those who have had at least two prior therapies if the patient cannot have (autoSCT) or multi-agent chemotherapy.[5]

The company state that 'for those who do not respond to BV the prognosis remains poor with little / no treatment options.'[1] There is no standard therapy after autoSCT and BV.[4] BV can be used as retreatment according to its licence but no specific recommendations have been made by NICE regarding retreatment.[5] Single or combination treatments including different chemotherapy regimens (some outside their marketing authorisation) may be used. This is the point in the clinical pathway at which pembrolizumab is aimed.

Pembrolizumab is therefore at least a third line treatment for people with relapsed or refractory cHL who have received autologous stem cell transplant (autoSCT) and brentuximab (BV) or BV when autoSCT is not a treatment option. For those who are suitable, pembrolizumab represents a bridge to allogeneic SCT (alloSCT), a potentially curative treatment.

ERG comment:

- The company's overview of current service provision is appropriate and relevant to the decision problem under consideration.
- Although not listed as a comparator in the NICE scope and not referenced in the CS, nivolumab has recently received approval from NICE for this condition. It is recommended 'as an option for treating relapsed or refractory classical Hodgkin lymphoma in adults after autologous stem cell transplant and treatment with brentuximab vedotin.'[6] Nivolumab is, however, not recommended for one of the populations in this appraisal (those who have received BV but who have not received an autoSCT).

3. CRITIQUE OF COMPANY'S DEFINITION OF DECISION PROBLEM

Table 3.1: Statement of the decision problem (as presented by the company)

	Final scope issued by NICE	Decision problem addressed in the company submission and rationale	ERG comments
Population	People with relapsed or refractory classical Hodgkin Lymphoma who have received: • autologous stem cell transplant and brentuximab vedotin • brentuximab vedotin when autologous stem cell transplant is not a treatment option.	As per final scope	This is in accordance with the scope.
Intervention	Pembrolizumab	As per final scope	This is in accordance with the scope.
Comparator(s)	• Single or combination chemotherapy including drugs such as gemcitabine, vinblastine and cisplatin • Best supportive care.	Standard of care as per Cheah et al. 2016) including: • Investigational agent • Gemcitabine • Bendamustine • Other alkylatory • BV retreatment • Platinum based • autoSCT • Other Cheah et al. 2016 reported outcome data for a mix of chemotherapy regimens and was preferred by the ERG in TA462. To separate individual regimens survival outcome data would not have been possible in the absence of individual patient level data and hence conservatively MSD have included all survival outcomes reported here.	Not in line with the final scope. The company has not provided separate data for comparators; instead a combined data set has been provided for multiple comparators, some of which are within the scope and others not. In TA462[6],"The committee concluded that the Cheah study was the best available evidence for standard of care and considered it appropriate for its decision-making, but overall the clinical effectiveness of nivolumab compared with standard of care was highly uncertain because the comparator data may not fully represent UK clinical practice."

	Final scope issued by NICE	Decision problem addressed in the company submission and rationale	ERG comments
Outcomes	The outcome measures to be considered include: • overall survival • progression-free survival • response rates • adverse effects of treatment • health-related quality of life	As per final scope, with the exception of long term overall survival data. The model structure utilised OS data from week 0-12 from KEYNOTE-087, response rates at week 12, PFS from week 12 onward and external literature OS sources for post alloSCT survival. At follow up (15.9 month), there were insufficient mortality events and median OS [redacted]. Hence all available data from KEYNOTE-087 has been utilised where possible.	Mostly in line with the final scope. However, survival data (OS and PFS) are immature. In addition, only two outcomes have been included in the indirect comparison: PFS and ORR.
Economic analysis	The reference case stipulates that the cost effectiveness of treatments should be expressed in terms of incremental cost per quality-adjusted life year. The reference case stipulates that the time horizon for estimating clinical and cost effectiveness should be sufficiently long to reflect any differences in costs or outcomes between the technologies being compared. Costs will be considered from an NHS and Personal Social Services perspective. The availability of any patient access schemes for the intervention or comparator technologies will be taken into account.	As per final scope	In line with the scope. However, a minor remark, the time horizon of 40 years was too short to capture the lifetime of all patients. A time horizon of 50 years, which was sufficiently long, was used in scenario analysis. Furthermore, Best Supportive Care was not presented as a comparator, with the exception of a scenario analysis. The company justified this citing a lack of data.
Subgroups to be considered	If the evidence allows, a scenario analysis including allogeneic stem cell transplant as a subsequent treatment after pembrolizumab or its comparators will be considered. This should reflect the	No response.	Mostly in line with the scope. Allogeneic stem cell transplant was incorporated into the company's base-case model as a subsequent treatment, reflecting the proportion of people who proceed to it

	Final scope issued by NICE	Decision problem addressed in the company submission and rationale	ERG comments
	proportion of people who proceed to allogeneic stem cell transplant after each treatment, as well as the costs and quality-adjusted life year benefits of the procedure.		after each treatment, as well as costs and quality-adjusted life year benefits of the procedure. A model without this option was not provided.
Special considerations including issues related to equity or equality	Not applicable.	Not applicable.	

Source: Table 1, Section B.1.1 of the CS.[1] and NICE FAD for TA462[6]

alloSCT = Allogeneic Stem Cell Transplant; ERG = Evidence Review Group; MSD = Merck Sharp and Dohme Ltd; NHS = National Health Service; OS = Overall Survival; TA = Technology Assessment.

3.1 Population

The population of this appraisal is in line with the scope. The main trial in the CS covers both cohorts of interest (people with relapsed or refractory cHL who have received autologous stem cell transplant and brentuximab and those who have received BV when autoSCT is not a treatment option).

However, only four out of 69 patients in cohort 1 and 10 out of 81 patients in cohort 2 of the intervention study (KEYNOTE-087) were from the UK. None of the patients in the comparator study (Cheah et al. 2016[7]) were from the UK.

The comparator study in this appraisal was also used in a previous appraisal (TA462[6]). NICE concluded in TA462 that "the comparator data may not fully represent UK clinical practice". If that is the case, then the results of the Matching-Adjusted Indirect Comparison (MAIC) in this appraisal are also not representative for UK clinical practice. This is because, the MAIC aims to generate the effect of pembrolizumab that would be observed in the Cheah trial population.[8]

3.2 Intervention

The intervention (pembrolizumab) is in line with the scope. Regulatory approval by the EMA for the indication considered within this submission was granted on the 2 May 2017. This stated that pembrolizumab as monotherapy is indicated for the treatment of adult patients with relapsed or refractory classical Hodgkin Lymphoma who have failed autoSCT and BV, or who are transplant-ineligible and have failed BV.

Pembrolizumab is a highly selective humanised monoclonal antibody against programmed death-1 (PD-1) receptor, which exerts dual ligand blockade of the PD-1 pathway, including PD-L1 and PD-L2, on antigen presenting tumour cells. By inhibiting the PD-1 receptor from binding to its ligands, pembrolizumab activates tumour-specific cytotoxic T lymphocytes in the tumour microenvironment and reactivates anti-tumour immunity.

The route of administration for pembrolizumab is IV infusion, over a 30-minute period. The anticipated licensed dosing regimen for patients with relapsed or refractory classical Hodgkin Lymphoma who have failed autoSCT and Brentuximab Vedotin (BV), or who are transplant ineligible and have failed BV is 200 mg every three weeks. Treatment with pembrolizumab continues until disease progression or unacceptable toxicity, whichever occurs first. The list price of pembrolizumab is £2,630 per 100 mg vial ([redacted]).[1]

3.3 Comparators

The description of the comparators in the NICE scope is as follows:

- Single or combination chemotherapy including drugs such as gemcitabine, vinblastine and cisplatin
- Best supportive care.

The company provides one study for the comparator. This is a retrospective USA database study published in 2016 by Cheah and colleagues in which patients received the following types of therapy: investigational agent(s), gemcitabine, bendamustine, any other alkylator, BV retreatment, platinum based treatment, autoSCT or alloSCT, or other treatment.[7] This is referred to in the CS as standard of care (SoC).

This comparator broadly matches the comparator described in the NICE scope: "Single or combination chemotherapy including drugs such as gemcitabine, vinblastine and cisplatin." However the ERG notes that there is some uncertainty about how well the Cheah study[7] which drew on data from patients treated

in the USA and which provides the base case comparator data, reflects the experience of patients treated in the UK. There is a lack of detail in the Cheah and colleagues' publication about the precise composition of the treatment regimens received by patients who had received ASCT and brentuximab vedotin. Many patients for whom outcome evaluations were available (28/67; 42%) were enrolled onto trial protocols and received what is described as 'Investigational agent', but there is no further detail about which therapies may have been classified under this heading. To find out whether treatments such as pembrolizumab were included among the 'Investigational agent' treatments, the ERG asked the company to clarify this (Clarification letter, Question A13).[9] The company replied that investigational agents did not included pembrolizumab although 'a couple of patients in the study received a PD-1 inhibitor.'[10] The company provided response rates results data excluding investigational agents.[10] The next most common regimens received by patients in the Cheah and colleagues study were gemcitabine-based (12/67; 18%) or bendamustine-based (11/67; 16%).

As reported in the ERG report for TA462, "gemcitabine regimens such as GDP (gemcitabine, dexamethasone, cisplatin) are commonly used in this patient population in the UK but platinum-containing regimens such as ESHAP (etoposide, methylprednisolone, cytarabine, cisplatin) and DHAP (dexamethasone, cytarabine, cisplatin) are also in common use. In the Cheah study 12/67 (18%) of patients with outcome evaluations received gemcitabine and just 4/67 (6%) of patients received platinum-based regimens."[11]

However, despite the uncertainty about how closely the experience of patients from the USA may match that of patients in the UK, the ERG is not aware of a more appropriate source of data for the comparator population.

Evidence for the clinical efficacy of best supportive care (BSC) is not presented within the clinical effectiveness section of the CS and in section 5 (cost effectiveness) of the CS, the company states that "Based on BCSH guidelines and clinician opinion, it is believed that use of BSC is minimal at this stage in the treatment pathway, as eligible patients are likely to receive therapy where feasible. As such, BSC has been applied within the model as a subsequent therapy in the base case analysis, with the composition derived from a recent NHL NICE Technology Appraisal (TA306).[12]" (CS, page 148)

In the economic model a scenario analysis was provided assessing the impact of BSC as a comparator. Due to lack of data informing the efficacy of BSC, in this scenario analysis, efficacy of BSC was assumed equivalent to that of Standard of Care (SoC).

3.4 Outcomes

The NICE final scope lists the following outcome measures:

- overall survival (OS)
- progression free survival (PFS)
- response rate (RR)
- adverse effects of treatment (AE)
- health-related quality of life (HRQoL)

These outcomes are reported in the CS. However, survival data (OS and PFS) are immature, and only two outcomes have been included in the indirect comparison: PFS and ORR.

3.5 Other relevant factors

According to the company a commercial access agreement (CAA) is in place with the Department of Health [redacted]) of the list price of pembrolizumab (CS, Table 4, page 31).

In addition, the company states that "no additional tests or investigations are required further to the usual tests undertaken in current clinical practice. No diagnostic test is required to identify the population for whom pembrolizumab is indicated and no particular administration for the technology is required." (CS, section 2.4, page 32).

Regarding the innovative nature of pembrolizumab, the company states that the US Food and Drug Administration (FDA) granted pembrolizumab Orphan Drug Designation for the treatment of HL, and Breakthrough Therapy Designation; in addition, the application received priority review status and accelerated approval.[13]

No equity or equality issues were specified in the final scope or identified by the company. The ERG is not aware of any issues related to equity or equality in the use of pembrolizumab in patients with relapsed or refractory classical Hodgkin lymphoma following (autoSCT and) brentuximab vedotin.

4. CLINICAL EFFECTIVENESS

4.1 Critique of the methods of review(s)

The company updated an existing systematic review to identify evidence on the use of pembrolizumab in classical Hodgkin Lymphoma. The review was designed to identify both clinical trials and observational studies and to inform both direct and indirect comparisons between the interventions relevant to the NICE scope. This section critiques the methods of the review including searching, inclusion criteria, data extraction, quality assessment and evidence synthesis.

4.1.1 Searches

The following paragraphs contain summaries and critiques of all searches related to clinical effectiveness presented in the company submission. The Canadian Agency for Drugs and Technologies in Health (CADTH) evidence based checklist for the Peer Review of Electronic Search Strategies (PRESS), was used to inform this critique.[14] The submission was checked against the Single Technology Appraisal (STA) specification for company/sponsor submission of evidence.[15] The ERG has presented only the major limitations of each search strategy in the report.

The company submission stated that systematic review searches were undertaken in October and December 2016, with an update in June 2017. Search strategies were reported in Appendix 2 of the CS for the following databases: Embase, MEDLINE, MEDLINE in-Process, Cochrane's CENTRAL database.

Additional searches of the following conference proceedings using the Northern Light database were reported: American Society of Clinical Oncology (ASCO) (2015-2016) and the American Society of Haematology (ASH) (2014-2016), as well as a manual search of the WHO International Clinical Trials Registry (WHO ICTRP) to identify ongoing trials.

Searches utilised study design filters based on the Scottish Intercollegiate Guidelines Network (SIGN) filters for RCTs and Observational Studies.[16]

ERG comment:

- The database searches were clearly structured and documented.
- The ERG was concerned that limiting the clinical effectiveness searches reported in Appendix 2 to English language only may have introduced language bias. Current best practice states that *"Whenever possible review authors should attempt to identify and assess for eligibility all possibly relevant reports of trials irrespective of language of publication"*.[17]
- Best practice outlined in the Cochrane handbook states that *"Reference lists in other reviews, guidelines, included (and excluded) studies and other related articles should be searched for additional studies"*.[18] However the ERG found no mention of reference checking within the report. It was unclear whether this was due to a reporting error or an omission within the SR process.
- Free text terms were used to search for relapsed/refractory in the search strategies for observational studies on the Embase, MEDLINE and MEDLINE in-Process databases. This facet could have been extended to a broader range of search terms i.e. resist$ or persist$ or return$ or reocur$ or reoccur$ or recurren$ or recidiv$ or regenerat$ and the inclusion of MeSH/Emtree terms such as relapse/. Given the low number of hits retrieved due to the addition of a facet for brentuximab vedotin, the inclusion of the line for relapsed/refractory terms may have been overly restrictive. However, this is unlikely to have greatly affected the overall recall of results.

4.1.2 Inclusion criteria

The eligibility criteria used in the search strategy for randomised controlled trials (RCTs) and non-RCTs is presented in Table 4.1.

The original review by the company was conducted in 2016 with an update in June 2017. The original inclusion criteria for the 2016 search strategy included a wider population and a longer list of interventions than the update. For the 2017 update search, the population was restricted to a population that was more in line with the final NICE scope (those who had disease progression during or after BV), and the interventions were defined in the same terms as those in the final scope: "Single or combination chemotherapy including drugs such as cisplatin, gemcitabine and vinblastine, and best supportive care". The updated review was designed to identify studies to inform both direct and indirect comparisons between interventions relevant to the NICE scope. The CS stated that two reviewers were involved in study selection with a third consulted in case of discrepancies.

Table 4.1: Eligibility criteria used in search strategy

	Description	
	Original SLR (Oct.19 and Dec. 2, 2016)	**Updated SLR (June 15 2017)**
Population	Adult cHL patients who either: failed to achieve a response to any line of therapy (refractory patients) or who have relapsed after ≥ 3 prior lines of therapy	Additional criteria added to restrict patients to those with disease progression during or after treatment with BV
Interventions	The following targeted drugs alone or as combinations with systemic chemotherapies: • Pembrolizumab • Brentuximab vedotin • Everolimus • Lenalidomide • Lucatumumab • Nivolumab • Ofatumumab • Panobinostat • Rituximab • Vorinostat The following systemic chemotherapies alone or in combinations: • Adriamycin • Bendamustine • Bleomycin • Carmustine • Cisplatin • Cyclophosphamide • Cytarabine • Dacarbazine • Etoposide • Gemcitabine • Ifosfamide • Mecholrethamine (Nitrogen mustard) • Melphalan • Mitoxantrone • Oxaliplatin • Procarbazine • Vinblastine • Vincristine • Vinorelbine Other treatments in combination with chemotherapies: • Dexamethasone • Methylprednisolone • Prednisone	Additional criteria were added to reflect only those interventions considered relevant to UK clinical practice: • Single or combination chemotherapy including drugs such as: o Cisplatin o Gemcitabine o Vinblastine • Best supportive care
Comparators	Any	Any

	Description	
	Original SLR (Oct.19 and Dec. 2, 2016)	**Updated SLR (June 15 2017)**
Outcomes	• Overall survival • Progression-free survival • Objective response • Complete response • Partial response • Treatment discontinuation due to AEs • Serious (grade 3 and above) AEs (not used for study selection)	No change
Study design	• Randomised controlled trials • Non-randomised controlled trials • Single arm trials • Retrospective and prospective controlled observational studies • Single group observational studies	No change
Source: CS, Table 6, page 43[1] AEs = adverse events; BV = brentuximab vedotin; cHL = classical Hodgkin Lymphoma		

ERG comment:

- The restriction of the updated systematic review to a population more in line with the NICE scope was appropriate.
- The original criteria for the 2016 systematic review included a longer list of interventions. For the 2017 update the interventions were defined in the same terms as those in the final scope: "Single or combination chemotherapy including drugs such as cisplatin, gemcitabine and vinblastine, and best supportive care". However, the phrase 'drugs such as' is rather vague and studies were excluded because the treatment 'did not reflect UK practice'; therefore, we asked the company to specify which interventions were included (Clarification letter, Question A2). The company responded by repeating the NICE scope: "MSD included comparators listed in the NICE final scope (March 2017) considered to represent UK clinical practice. This comprises: single or combination chemotherapy including drugs such as gemcitabine, vinblastine, and cisplatin and best supportive care."[10] The impact of this is discussed in Section 4.2.1 of this report.

4.1.3 Critique of data extraction

The CS stated that two investigators extracted data independently from the included studies. Any discrepancies between data extractions were resolved by involving a third reviewer and coming to a consensus.

ERG comment:

- Data extraction appears to have been conducted appropriately.

4.1.4 Quality assessment

It appears that two investigators assessed study quality independently. Any discrepancies between assessments were resolved by involving a third reviewer and coming to a consensus. The tool used was the Newcastle-Ottawa Scale covering issues related to selection bias and assessment of outcomes.

ERG comment:

- Study quality was assessed appropriately. Results of the quality assessment by the company and the ERG of the KEYNOTE-087 trial are outlined in Section 4.2.2.4 of this report. The limitations of single-arm studies are also outlined in Section 4.2 of this report.

4.1.5 Evidence synthesis

No trials directly comparing pembrolizumab with a comparator of interest were identified therefore a meta-analysis of the direct evidence could not be performed. The company described the results of the KEYNOTE-087 single arm trial. A retrospective observational study (Cheah et al. 2016[7]) was identified from searches of the literature and used as a comparator in naïve comparisons and matched adjusted indirect comparison (MAIC). This analysis and its results are described more fully in Section 4.4 of this report.

ERG comment:

- The ERG agrees that no direct meta-analysis was possible given that only one single arm study of pembrolizumab was identified (KEYNOTE-087).

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

4.2.1 Overview of the evidence in the submission

No relevant randomised controlled trials (RCTs) of pembrolizumab were identified by the company. The CS was based on one ongoing single arm phase II trial (KEYNOTE-087). KEYNOTE-087 will be discussed in detail in this section of the report.

The submission briefly mentions a phase 1b trial of pembrolizumab (KEYNOTE-013). However, this study did not correspond to the EMA licensing for the dosing of pembrolizumab and was used as supporting evidence for safety only, therefore will only be briefly mentioned in Section 4.2.3 of this report. The company also provides details of a clinician survey to support understanding of UK clinical practice. This survey is also briefly discussed in Section 4.2.3 of this report.

Two further trials were mentioned as being ongoing: KEYNOTE-204 and NCT03077828. These studies are discussed in Section 4.2.4 of this report.

ERG comment:

- The ERG was provided with a list of excluded studies. The company stated in response to clarification that 'MSD included comparators listed in the NICE final scope (March 2017) considered to represent UK clinical practice. This comprises: single or combination chemotherapy including drugs such as gemcitabine, vinblastine, and cisplatin and best supportive care.'[10] The ERG checked the list of studies excluded based on intervention and concluded that no comparative studies had been inappropriately excluded.
- A small number of studies of nivolumab were identified and excluded. Although not listed as a comparator in the NICE scope and not referenced in the CS, nivolumab has recently received approval from NICE for this condition. It is recommended 'as an option for treating relapsed or refractory classical Hodgkin lymphoma in adults after autologous stem cell transplant and treatment with brentuximab vedotin.'[6] Nivolumab is, however, not recommended for one of the populations in this appraisal (those who have received BV but who have not received an autoSCT).

- Bendamustine has also been investigated in small observational studies which were excluded from the review. The ERG did not believe these studies were suitable comparator studies and agreed that they should be excluded.[19, 20]
- The ERG notes that the evidence for pembrolizumab is based on one single arm, ongoing trial.

4.2.2 KEYNOTE-087

4.2.2.1 Methodology of KEYNOTE-087

KEYNOTE-087 is a phase II, multicentre, single arm trial of pembrolizumab in adult patients with RRcHL. See Table 4.2.

Table 4.2: Methodology of the KEYNOTE-087 trial

PICOS	Details		
Population	Patients ≥ 18 with relapsed[a] or refractory[b] de novo classical Hodgkin Lymphoma		
	Measurable disease defined as ≥ 1 lesion accurately measured in ≥ 2 dimensions with spiral CT. Minimum measurement > 15mm in the longest diameter or > 10mm in the short axis.		
	ECOG Performance Scale 0 or 1		
	Cohort 1 (n = 69)	*Cohort 2 (n = 81)*	*Cohort 3[c]*
	Have failed to achieve a response or have progressed after autoSCT.	Were unable to achieve a CR or a PR to salvage chemotherapy and did not receive autoSCT.	Have failed to achieve a response or have progressed after autoSCT
	Patients must have relapsed after treatment with or failed to respond to BV post autoSCT.	Patients must have relapsed after treatment with or failed to respond to BV	Patients have not received BV post autoSCT.
Setting	Three study sites in the UK[d], 23 elsewhere in Europe, 11 in the USA, seven in Japan, four in Israel, two in Australia and one in Canada		
Intervention	200mg pembrolizumab as 30 min IV infusion every three weeks in the outpatient setting		
Outcomes[e]	*Primary*	*Secondary*	
	Overall response rate (ORR) defined as the proportion of patients who have complete remission (CR) or partial remission (PR) using IWG response criteria assessed by CT / PET at any time during the study as determined by blinded, independent central review (BICR).	ORR using IWG criteria at any time during the study as determined by investigator	
	Safety and tolerability (including adverse events and serious adverse events)	ORR using 5-point scale according to the Lugano classification as determined by BICR	
		Progression-free survival (PFS) and duration of response (DOR) by BICR and by investigator according to the IWG criteria	

PICOS	Details	
		Overall survival
Study design	Phase II single arm, open label trial	

Source: Section 4.3.1 of the CS
Footnote: a) Disease progression after most recent therapy; b) failure to achieve CR or PR to most recent therapy; c) Not relevant to this appraisal; d) 14 patients were from the UK (Cohort 1, n = 4; Cohort 2, n = 10); e) The trial also listed exploratory outcomes including an assessment of ORR, CRR, PFS and DOR for patients who continue treatment with pembrolizumab beyond documented progression and an assessment of health-related quality of life.
autoSCT = Autologous Stem Cell Transplant; BICR = Blinded independent central radiologists; BV = Brentuximab Vedotin; CR = Complete response; ECOG = Eastern Cooperative Oncology Group; IV = intravenous; IWG = International Working Group; ORR = Objective Response Rate; PR = Partial response

The trial has three cohorts. Cohort 1 includes patients who have failed to achieve a response or who have progressed after autoSCT and have relapsed after treatment with or have failed to respond to BV post autoSCT. Cohort 2 comprises patients, most of whom were unable to achieve CR or PR to salvage chemotherapy and did not receive autoSCT, and have relapsed after treatment with or failed to respond to BV. Cohort 3 includes patients who have failed to respond to, or have progressed after autoSCT and have not received BV post autoSCT (see Table 4.2). Cohort 3 is not relevant to this submission so effectiveness results are not presented for this cohort in this report.

A number of patient exclusion criteria were outlined in the CS. Most relevant are that patients who had undergone prior alloSCT within the last five years were excluded. Patients who had a transplant greater than five years ago were eligible provided there were no symptoms of graft vs. host disease. A further exclusion criterion was that patients should not have received prior therapy with an anti-PD-1, anti-PD-L1, anti-PD-L2, anti-CD-137 or anticytotoxic T-lymphocyte associated antigen-4 antibody (including ipilimumab or any other antibody or drug specifically targeting T-cell co-stimulation or checkpoint pathways).

KEYNOTE-087 has a total of 210 participants of whom 150 are relevant to this submission (69 cohort 1, 81 cohort 2). It is a multinational trial including three sites in the UK. The CS further detailed that four patients in cohort 1 and 10 in cohort 2 were from the UK.

As a single arm, open label trial, treatment was known to both investigators and patients. Patients received 200 mg pembrolizumab as 30 min IV infusion every three weeks in the outpatient setting. Neither dose escalation nor dose reduction of pembrolizumab was permitted in the trial. Dose modification due to adverse events (both serious and non-serious) was permitted. All concomitant permitted medications received within 28 days before the first dose of trial treatment and 30 days after the last dose of trial treatment were recorded.

Disease response assessments were planned for every 12 weeks until documented disease progression, the start of a new anti-cancer treatment, withdrawal of consent, death or the end of the study, whichever occurred first. Bone marrow biopsies were collected to confirm complete remission (in patients who had bone marrow involvement) or if clinically indicated. Where a patient showed progressive disease pembrolizumab could be continued at the discretion of the principal investigator (PI) until the next disease response assessment provided their clinical condition was stable. Imaging should have occurred at any time where there was clinical suspicion of progression. Patients who experienced a complete or partial response or had stable disease were able to remain on treatment for up to two years (approximately 37 administrations) or until unacceptable toxicity or progression. Patients who attained

a complete response could stop pembrolizumab after a minimum of 24 weeks of treatment with at least two doses since initial confirmation of CR. Patients who later experienced disease progression could be retreated with pembrolizumab at the same dose and schedule as at the time of initial discontinuation if no cancer treatment had been administered since the last dose of pembrolizumab. [redacted]

The primary outcome was best overall response rate (best ORR or BOR); ORR is defined as the proportion of patients who have complete remission (CR) or partial remission (PR) using International Working Group (IWG) response criteria assessed by CT/PET at any time during the study as determined by blinded, independent central review (BICR), and the Best Overall Response (BOR) is the best response recorded from the start of the study treatment until the disease progression/recurrence. Progression free survival and overall survival were assessed as secondary outcomes.

Health-related quality of life was also evaluated as an exploratory outcome. Assessments were made from baseline using the European Organisation for Research and Treatment of Cancer (EORTC) Quality of Life (QoL) Questionnaire C30 (QLQ-C30) and European Quality of Life Five Dimensions Questionnaire (EuroQoL, EQ-5D).

ERG comments:

- The most important methodological aspects to note are that although the trial was well conducted, it represents a low level of evidence. It is a phase II, single arm, non-comparative study which by its design has serious limitations. We cannot know whether the outcomes observed are a true reflection of the intervention as the role of natural history and baseline characteristics is not taken into account. This is in contrast to a well-conducted randomised trial where bias is minimised and we can be confident that outcomes we observe are due to differences between the interventions evaluated.
- As a single-arm, open-label trial the intervention is known to participants, clinicians and assessors. Knowledge of interventions can lead to bias in delivering interventions and reporting outcomes.
- The trial gives a maximum two years of outcome data on patients. Outcomes are relevant but the primary outcome is objective response rate rather than the longer-term outcomes of PFS and OS which are evaluated as secondary outcomes.
- Cohorts 1 and 2 of the trial are relevant to the decision problem in the NICE scope. Inclusion and exclusion criteria in terms of population appear to be appropriate.
- Although the trial is multinational it only has 150 relevant participants so the evidence base for this appraisal is small. However, the population matching the scope of this appraisal is in itself small so conducting a larger trial would be challenging.
- A small number of patients were from the UK (14) so the trial may not totally reflect the UK population and setting. However, once again the population from which to draw participants is small.
- In clinical practice, for those who are suitable, pembrolizumab represents a bridge to alloSCT, a potentially curative treatment. However, the company submission stated that 'KEYNOTE-087 was not designed as a 'bridging' study, therefore the uptake of alloSCT was very low overall across cohorts 1 and 2.'[1] The company further stated that 'the use of stem cell transplant would have been at the discretion of the treating physician on a per patient basis.'[10] The company clarified that [redacted] of the KEYNOTE-087 population had an alloSCT following pembrolizumab ([redacted] from cohort 1 and [redacted] from cohort 2).[10] However it was noted

that [redacted] patients in cohort 1 and 2 in the UK were transplanted with allogeneic stem cells respectively.[1]

4.2.2.2 Statistical analysis of KEYNOTE-087

The primary hypothesis of this study was that i.v. administration of pembrolizumab would reach an ORR of greater than 20% in each of the three cohorts using IWG criteria by independent review committee. The selection of 20% as a control rate was based partly on the published literature prior to the approval of BV and downgraded to take account that this patient group have failed treatment with BV. Enrolment of 60 patients per cohort was required to have 93% power at a one-sided 2.5% α level to detect a 40% or higher ORR for pembrolizumab compared to a fixed control rate of 20% using an exact binomial test. Each cohort was analysed separately and also as a pooled group. However, only the results for cohorts 1 and 2 are presented in this report. The company stated that '*No additional multiplicity adjustment was required because each cohort was evaluated independently*.'[1]

Final analysis was to be conducted for each cohort when the last participant reached the Week 12 response assessment or discontinued study therapy. Results are presented as a percentage with the exact 95% two-sided CI (Clopper-Pearson method). An exact binomial test was used to obtain a one-sided p-value for comparing the observed ORR to the control value of 20% (null hypothesis $p \leq 0.20$ vs. alternative hypothesis $p > 0.20$) for each cohort. The analysis of the primary endpoints used the All Subjects as Treated (ASaT) population (those who had received at least one dose of medication). Supportive analyses were also conducted using the full analysis set (FAS) which was all patients who received at least one dose of study medication, had a baseline disease assessment and either a post-baseline disease assessment or who discontinued the trial due to progressive disease/drug related AE.

Time to event outcomes (response duration, PFS and OS) were summarised by the median time to event with 95% CI using the Kaplan-Meier method. The percentage surviving at different time points (3, 6, 9 and 12 months for PFS) and for OS (6, 9, 12 and 15 months) were also obtained using the Kaplan-Meier method,

The ASaT population was also used for the analysis of safety. Additionally, at least one laboratory or vital sign measurement obtained subsequent to at least one dose of trial treatment was required for inclusion in the analysis of each specific safety parameter.

Table 4.3 gives an overview of the main analyses undertaken in KEYNOTE-087.

Table 4.3: Efficacy analysis of primary and secondary endpoints in KEYNOTE-087

Endpoint / Variable	Statistical method	Analysis population	Missing data approach
Primary outcome			
Overall response rate IWG criteria (2007) • Central review	Exact test of binomial parameter; 2-sided 95% exact CI	ASaT / FAS	Participants with missing data are considered non-responders
Secondary outcomes			
Overall response rate IWG criteria (2007) • Study site Lugano criteria (2014) • Central review	Point estimate; 2-sided 95% exact CI	ASaT / FAS	Participants with missing data are considered non-responders

Endpoint / Variable	Statistical method	Analysis population	Missing data approach
Complete remission rate IWG criteria (2007) • Central review • Study site Lugano criteria (2014) • Central review	Point estimate; 2-sided 95% exact CI	ASaT / FAS	Participants with missing data are considered non-responders
Progression-free survival IWG criteria (2007) • Central review • Study site	Summary statistics using Kaplan-Meier method	ASaT / FAS	Censored at last assessment
Duration of response IWG criteria (2007) • Central review • Study site	Summary statistics using Kaplan-Meier method	All responders	Non-responders are excluded in analysis
Overall survival	Summary statistics using Kaplan-Meier method	ASaT / FAS	Censored at last assessment
Source: CS, Table 9, page 62. ASaT = All Subjects as Treated; CI = Confidence Interval; FAS = Full Analysis Set; IWG = International Working Group.			

ERG comment:

- The ERG has no concerns about the design or statistical analyses of the KEYNOTE-087 trial. It was a non-comparative single-arm trial and the sample size calculation and analysis methods are appropriate.

4.2.2.3 Participants in the KEYNOTE-087 trial

A total of 210 patients were enrolled in the KEYNOTE-087 trial of which 69 formed cohort 1 (who had failed to achieve a response or had progressed after autoSCT) and 81 formed cohort 2. In response to clarification the company stated that the majority of cohort 2 did not qualify for an autoSCT [redacted]. This information was not routinely gathered but the company stated that a small number of participants did not receive autoSCT for a variety of reasons [redacted].[10]

All patients in both cohorts had relapsed or refractory disease and all had used BV as per the inclusion criteria for the trial. Cohort 3 (the remaining 60 patients) are not relevant to this appraisal. Patient characteristics for cohorts 1 and 2 are reported in Table 4.4.

Table 4.4: Patient characteristics in the KEYNOTE-087 trial

	Cohort 1 (n = 69) n (%)	Cohort 2 (n = 81) n (%)
Gender		
Male	36 (52.2)	43 (53.1)
Female	33 (47.8)	38 (46.9)
Age Years		
<65	69 (100)	66 (81.5)

	Cohort 1 (n = 69) n (%)	**Cohort 2 (n = 81) n (%)**
≥ 65	0	15 (18.5)
Mean (SD)	37.0 (10.9)	42.3 (17.4)
Median	34.0	40
Range	19 to 64	20 to 76
Race		
American Indian or Alaska native	0	1 (1.2)
Asian	7 (10.1)	4 (4.9)
Black or African American	2 (2.9)	2 (2.5)
Missing	1 (1.4)	1 (1.2)
Multi-racial	2 (2.9)	0
White	57 (82.6)	73 (90.1)
Disease subtype		
CHL – nodular sclerosis	55 (79.7)	65 (80.2)
CHL – mixed cellularity	9 (13.0)	10 (12.3)
CHL – lymphocyte rich	4 (5.8)	1 (1.2)
CHL – lymphocyte depleted	0	4 (4.9)
Missing	1 (1.4)	1 (1.2)
ECOG performance status		
0	29 (42.0)	44 (54.3)
1	39 (56.5)	37 (45.7)
2	1 (1.4)	0
Prior lines of therapy		
≥ 3	68 (98.6)	78 (96.3)
< 3	1 (1.4)	3 (3.7)
Mean (SD)	4.5 (1.7)	4.0 (1.7)
Median	4.0	4.0
Range	2 to 12	1 to 11
Time of relapse since SCT failure (months)		
≥ 12	37 (53.6)	0
< 12	32 (46.4)	0
Mean (SD)	60.2 (39.6)	NA
Median	12.6	NA
Range	2.5 to 247.9	NA
Prior radiation		
Yes	31 (44.9)	21 (25.9)
No	38 (55.1)	60 (74.1)
Bulky lymphadenopathy		
Yes	5 (7.2)	12 (14.8)

	Cohort 1 (n = 69) n (%)	Cohort 2 (n = 81) n (%)
No	64 (92.8)	69 (85.2)
Baseline B symptoms		
Yes	22 (31.9)	26 (32.1)
No	47 (68.1)	55 (67.9)
Baseline bone marrow involvement		
Yes	3 (4.3)	5 (6.2)
No	66 (95.7)	75 (92.6)
Missing	0	1 (1.2)

Source: Table 11 of the CS (abbreviated)
CHL = Classical Hodgkin Lymphoma; ECOG = Eastern Cooperative Oncology Group; SCT = stem cell transplant; SD = standard deviation

Both cohorts had slightly more male than female participants (52.2% in cohort 1 and 53.1% in cohort 2). Most participants across the cohorts were white (82.6% in cohort 1 and 90.1% in cohort 2). Although both cohorts had a wide age range (cohort 1: 19 to 64, cohort 2: 20 to 76) all of the participants in cohort 1 and 85.1% of the participants in cohort 2 were under 65 years of age. The most common disease subtype was cHL – nodular sclerosis (cohort 1: 79.7%, cohort 2: 80.2%). All patients except one in cohort 1 had an ECOG score of 0 or 1. Approximately a third of patients across the cohorts had B symptoms (cohort 1: 31.9%, cohort 2: 32.1%). A small number had bone marrow involvement (cohort 1: 4.3%, cohort 2: 6.2%).

The mean time of relapse since autoSCT was 60.2 months. Both cohorts were heavily pre-treated. In cohort 1 [redacted] of patients had received at least three lines of therapy (range 2 to 12). In cohort 2 [redacted] had received at least three lines of therapy (range 1 to 11). Participants in cohort 1 had a median of [redacted] days since the last dose of BV (Range [redacted] days) whilst cohort 2 had a median of [redacted] days since their last dose (range [redacted] days).

ERG comment:

- There is a peak in incidence of cHL in older males and females (75 to 79 for men and 70 to 74 for women) but no patients in cohort 1 are 65 or over 65. In cohort 2 18.5% of patients are 65 or over. At least in cohort 1 older patients are underrepresented in KEYNOTE-087.
- Advisers to the company stated that typically patients within the UK would have received between three and four prior lines of therapy, including BV, before starting treatment with a PD-L1. In the trial only [redacted] patient in cohort 1[redacted] and [redacted] in cohort 2 ([redacted] had received fewer than three therapies so in this respect is applicable to UK practice. However, it should be noted that [redacted] of cohort 1 ([redacted]) had received five or more therapies. In cohort 2 [redacted] ([redacted]) had received five or more therapies. The population of the trial could, therefore, be more heavily treated than in UK practice.

4.2.2.4 Quality assessment of the KEYNOTE-087 trial

The results of the company's and the ERG's assessment of KEYNOTE-087 are shown in Table 4.5. It should be noted that not all of the questions in the tool are applicable to a single-arm study.

Table 4.5: Quality assessment of the KEYNOTE-087 trial

	CS evaluation	ERG evaluation	ERG comment
Selection bias			
Representativeness of cohort	*	*	Representative of the cHL population but may not be representative of the UK population
Selection of non-exposed cohort	NA	NA	
Ascertainment of exposure	*	*	Assessment was made of number of patients who received at least one dose of treatment
Outcome of interest	*	*	Presence of the outcome of interest was assessed before exposure to the intervention.
Comparability of cohorts	NA	NA	
Outcome bias			
Outcome assessment	*	*	Outcomes were evaluated by an independent review committee (IRC).
Adequate duration of follow-up			Median follow up time was 15.9 months. This was adequate for ORR but not for PFS and OS.
Adequate follow-up of cohort	*	*	Explanations were provided regarding missing data or loss to follow up.
Source: CS, Table 12, page 68 CS = company submission; ERG = evidence review group; NA = non-applicable			

ERG comments:

- The most important methodological aspect to note is that although the trial was well conducted, it represents a low level of evidence. It is a phase II, single arm, non-comparative trial which by its design has serious limitations. We cannot know whether the outcomes observed are a true reflection of the intervention. The role of natural history and baseline characteristics is not taken into account.
- The study had an adequate follow-up (median 15.9 months) for the main outcome evaluated (ORR defined as the proportion of patients who have complete remission (CR) or partial remission (PR)). However median progression free survival was immature and [redacted]

4.2.2.5 Main efficacy results of the KEYNOTE-087 trial

At the 21 March 2017 data cut off [redacted] of cohort 1 patients and [redacted] of cohort 2 patients remained on treatment. Table 4.6 gives the current status of the patients in the KEYNOTE-087 trial.

Table 4.6: Patient status in the KEYNOTE-087 trial

Patient Status	Cohort 1 (n = 69) n (%)	Cohort 2 (n = 81) n (%)
Started	[redacted]	[redacted]
Discontinued	[redacted]	[redacted]
Adverse event	[redacted]	[redacted]
Bone marrow transplant	[redacted]	[redacted]
Clinical progression	[redacted]	[redacted]
Complete response	[redacted]	[redacted]
Death	[redacted]	[redacted]
Lost to follow-up	[redacted]	[redacted]
Physicians Decision	[redacted]	[redacted]
Pregnancy	[redacted]	[redacted]
Progressive disease	[redacted]	[redacted]
Withdrawal by subject	[redacted]	[redacted]
Treatment on-going	[redacted]	[redacted]
Source: CS, Table 10, page 64		

Table 4.7 presents the main efficacy data for the trial. The company confirmed that these were the latest efficacy data available.

Table 4.7: Summary efficacy results of the KEYNOTE-087 trial

Outcome[a]	Results[b]	
	Cohort 1 N = 69	Cohort 2 N = 81
Overall survival		
Death n (%)	[redacted]	[redacted]
Median (95% CI) months[c]	[redacted]	[redacted]
OS at 12 months % (95% CI)[c]	[redacted]	[redacted]
Progression-free survival		
Median (95% CI) months[c]	16.7 (11.2 to NR)	11.1 (7.6 to 13.7)
PFS at 12 months % (95% CI) [c]	[redacted]	[redacted]
Response rates		
ORR n (%)[d]	52 (75.4)	54 (66.7)
CR n (%)	19 (27.5)	20 (24.7)
PR n (%)	33 (47.8)	34 (42)
SD n (%)	[redacted]	[redacted]
PD n (%)	[redacted]	[redacted]
No assessment n (%)	[redacted]	[redacted]
Time to response Median (range) months[c]	[redacted]	[redacted]
Duration of response Median (range) months[c]	[redacted]	[redacted]

Outcome[a]	Results[b]	
	Cohort 1 N = 69	Cohort 2 N = 81
Source: CS, Section 4.7, tables 14 and 15 Footnote: a) as per the NICE scope; b) 21 March 2017 unless otherwise stated. Median follow-up 15.9 months (range 1.0 to 20.9 months); c) From product-limit (Kaplan-Meier) method for censored data; d) assessed by BICR using IWG criteria CI = confidence interval; ORR = overall response rate; OS = overall survival; PD = progressive disease; PFS = progression free survival; PR = partial response; SD = stable disease		

Overall response rate (the primary outcome as assessed by the independent committee using IWG criteria) was 75.4% in cohort 1 and 66.7% in cohort 2. In Cohort 1 27.5% of patients had a complete response and in cohort 2 this figure was 24.7%. Median time to response was [redacted] and [redacted] respectively. However median duration of response was [redacted] in cohort 1 and was [redacted] months in cohort 2.

Median PFS in cohort 1 as assessed by independent committee was 16.7 months (11.2 to NR). In cohort 2 it was 11.1 months (7.6 to 13.7).

Median OS was [redacted]. At 12 months survival was [redacted] in cohort 1 and [redacted] in cohort 2.

ERG comment:

- As stated above, the trial was long enough to show the benefit of pembrolizumab on overall response rates including both CR and PR. However, PFS and OS data are not fully mature.

4.2.2.6 Post-hoc analyses of the KEYNOTE-087 trial

The company conducted post-hoc analyses of response to inform the naïve indirect treatment comparison and the Matched Adjusted Indirect Comparison (MAIC) (discussed in Section 4.3). The main difference between this post-hoc analysis of response and the primary analysis of response, referred to as 'best' response rate, is that response was determined at a single time point for each patient i.e. 12 weeks as opposed to any time point up to the point of progression. Data in the form of the proportion who respond by a specific time point was required in order to apportion patients that were progression-free into CR, PR or SD in the cost-effectiveness model. The table below shows the response rates at week 12.

Table 4.8: 12-week response results in the KEYNOTE-087 trial

Outcome[a]	Results[b]	
	Cohort 1 N = 69	Cohort 2 N = 81
Response rates		
ORR n (%)[d]	[redacted]	[redacted]
CR n (%)	[redacted]	[redacted]
PR n (%)	[redacted]	[redacted]
Stable disease (SD) n (%)	[redacted]	[redacted]
Progressive disease (PD) n (%)	[redacted]	[redacted]

Outcome[a]	Results[b]	
	Cohort 1 N = 69	Cohort 2 N = 81
No assessment n (%)	[redacted]	[redacted]

Source: Section 4.8 of the CS
Footnote: a) as per the NICE scope; b) 21 March 2017 unless otherwise stated. Median follow-up 15.9 months (range 1.0 to 20.9 months); c) From product-limit (Kaplan-Meier) method for censored data; d) assessed by BICR using IWG criteria
ORR = overall response rate; CR = complete response; PR = partial response; SD = stable disease; PD = progressive disease

ERG comment:

- The ERG noted that overall response rates were lower at 12 weeks than over the course of the trial ([redacted]

4.2.2.7 Safety results of the KEYNOTE-087 trial

Safety results from KEYNOTE-087 are presented from the data cut-off of 25 September 2016. The company confirmed that these were the most recent safety data available. All enrolled patients received at least one dose of study treatment. The median time on treatment was [redacted] days for cohort 1 and [redacted] days for cohort 2. Cohort 1 had a mean number of [redacted] administrations whilst cohort 2 had a mean of [redacted].

Table 4.9 gives an overview of the numbers and percentages of patients who had an AE up to 30 days and serious AEs up to 90 days after the last dose of study medication.

Table 4.9: Overview of adverse events in the KEYNOTE-087 trial

Database Cut-off Date: 25 Sep 2016	Cohort 1 (n = 69)	Cohort 2 (n = 81)
Patients in population	**n (%)**	**n (%)**
with ≥ 1 AE	[redacted]	[redacted]
with no AE	[redacted]	[redacted]
with drug-related[a] adverse events	[redacted]	[redacted]
with toxicity grade 3-5[b] adverse events	[redacted]	[redacted]
with toxicity grade 3-5 drug-related adverse events	[redacted]	[redacted]
with non-serious adverse events	[redacted]	[redacted]
with serious adverse events	[redacted]	[redacted]
with serious drug-related adverse events	[redacted]	[redacted]
who died	[redacted]	[redacted]
who died due to a drug-related adverse event	[redacted]	[redacted]
Discontinued[d] due to an adverse event	[redacted]	[redacted]
discontinued due to a drug-related adverse event	[redacted]	[redacted]
discontinued due to a serious adverse event	[redacted]	[redacted]
discontinued due to a serious drug-related adverse event	[redacted]	[redacted]

Database Cut-off Date: 25 Sep 2016	Cohort 1 (n = 69)	Cohort 2 (n = 81)
Source: Table 39 of the CS AE = adverse event		

In cohort 1 [redacted] of patients had one or more adverse events. In cohort 2 [redacted] of patients had one or more adverse events. The company noted that most AEs were low grade ([redacted] and [redacted] grades 3 to 5 in cohort 1 and 2 respectively). In cohort 1 [redacted] of AEs were classed as serious and with approximately half of these drug-related. Similarly [redacted]% of cohort 2 experienced serious AEs of which approximately a quarter were serious. [redacted]. A small number of patients in both cohorts discontinued due to an adverse event ([redacted] in cohort 1, [redacted] in cohort 2).

The most common adverse events were pyrexia ([redacted]), cough ([redacted]) fatigue ([redacted]), diarrhoea ([redacted]) and vomiting ([redacted]).

[redacted] of AEs were deemed to be drug-related in cohort 1, and [redacted] of AEs were deemed to be drug-related in cohort 2. The most common drug-related AEs were hypothyroidism [redacted]) pyrexia ([redacted]), fatigue ([redacted]), rash ([redacted]) diarrhoea ([redacted]) and headache [redacted]).

Table 4.10 lists the drug-related serious adverse events by category in KEYNOTE-087. A SAE was defined as any AE that occurred during the use of pembrolizumab that resulted in: death, was life threatening, resulted in persistent or significant disability/incapacity, resulted in, or prolonged, an existing in-patient hospitalisation, was a congenital anomaly/birth defect, or was considered as another important medical event.

Table 4.10: Drug-related serious adverse events in the KEYNOTE-087 trial

Database Cut-off Date: 25 Sep 2016 up to 90 days after last dose	Cohort 1 (n = 69)	Cohort 2 (n = 81)
Patients in population[a]	**n (%)**	**n (%)**
One or more serious AE	[redacted]	[redacted]
Cardiac disorders	[redacted]	[redacted]
Myocarditis	[redacted]	[redacted]
Pericarditis	[redacted]	[redacted]
Immune system disorders	[redacted]	[redacted]
Cytokine release syndrome	[redacted]	[redacted]
Infections and infestations	[redacted]	[redacted]
Herpes simplex	[redacted]	[redacted]
Herpes zoster	[redacted]	[redacted]
Myelitis	[redacted]	[redacted]
Injury, poisoning and procedural complications	[redacted]	[redacted]
Infusion-related reaction	[redacted]	[redacted]
Musculoskeletal and connective tissue disorders	[redacted]	[redacted]
Myositis	[redacted]	[redacted]

Database Cut-off Date: 25 Sep 2016 up to 90 days after last dose	Cohort 1 (n = 69)	Cohort 2 (n = 81)
Respiratory, thoracic and mediastinal disorders	[redacted]	[redacted]
Dyspnoea	[redacted]	[redacted]
Pneumonitis	[redacted]	[redacted]
Source: Table 47 of the CS Footnote: a) Adverse events appear in this table if > 0 in Cohort 1 or 2. AE = adverse event		

ERG comment:

- Patients will need to be informed of the adverse events to make an informed decision on treatment. The percentage of drug-related events is high ([redacted] and [redacted] in cohorts 1 and 2 respectively). [redacted] of adverse events were grades 3 to 5 and in cohort 1 [redacted] were serious, in cohort 2 [redacted]. Given that nivolumab is now available for patients in cohort 1 it will be important to compare their adverse event profile.

4.2.3 Supporting evidence

Lafferty et al.[21]

The company used a study by Lafferty et al to provide data for the economic model of this appraisal.[21] Exact details of which data were used is discussed in the cost effectiveness section of this report. The study was not described in full in the CS and is only available as an abstract.

Briefly, the retrospective study evaluates 13 patients with HL who underwent alloSCT between 2008 and 2015. The population is described as being heavily pre-treated and all patients had received at least three lines of chemotherapy. Eight of 13 (62%) had undergone autoSCT prior to alloSCT. It was not stated if patients had received BV. Median age of the participants was 33. At the time of transplant 11 patients were in partial remission and two in complete remission. Donors were matched sibling (six patients), matched unrelated volunteer (six patients) and double cord stem cell transplant (one).

Median length of follow up in survivors was 424 days. At one year OS was 69% and PFS 54%. The four deaths in the first year were due to respiratory syncytial virus (RSV) infection, air embolism, acute graft versus host disease (GVHD). Relapse or progression post-transplant occurred in three patients (23%) all within one year. Acute GVHD developed in eight (62%) of patients and was grade II to IV in five (38%).

ERG comment:

- This study is relevant to the UK and was used in a previous appraisal (TA462).
- The study was available in abstract form only so could not be fully quality assessed. However, it is clear that as a source of data for the model, the study is very limited. It is a small, retrospective case series from a single centre in the UK. The care provided may not be typical of the general UK setting. The 13 patients may not fully reflect the characteristics of patients seen elsewhere in clinical practice. Older patients are not represented in this sample. It is unclear if all patients had received BV as per the population in this appraisal. There is no comparison of the outcome between those receiving alloSCT and those not. The role of natural history cannot be ascertained. The small numbers of patients mean that these results cannot be extrapolated to larger samples.

Clinician Survey

Due to the paucity of data available on standard of care for this patient group, the company commissioned a clinician survey to support understanding of UK clinical practice. Specifically, the survey aimed to determine UK clinical practice for the treatment of patients with RRcHL, to consider the treatment pathway and eligibility of patients with RRcHL following standard of care and to assess the validity of the Cheah et al.[7] study in relation to UK practice and the Lafferty et al.[21] study in relation to rates of alloSCT and outcomes after alloSCT in patients who have received standard of care in the relapsed/refractory setting.

The questionnaire was made available via a website and was completed by 16 clinicians (12 from England, one from Wales and three from Scotland). Respondents were either haematologists or haematological oncologists. The average number of patients seen by a clinician matching cohort 1 (failed autoSCT and BV) was four patients annually. The average number seen matching cohort 2 (ineligible for autoSCT and failed BV) was three patients annually. Three of 16 clinicians had experience of using PD-1s in cHL.

Clinicians noted that both cohorts of patients would receive standard of care for approximately 12 weeks. They considered that only a minority of patients on standard care would proceed to allogenic SCT when ineligible for autologous SCT and having failed BV (cohort 2 equivalent). This was estimated as 17% of those gaining a CR and 13% of those gaining a partial response. A CR to standard care was estimated as 12% of patients and a partial response to standard care was estimated from 19% of patients. This was in contrast to a cohort 1 equivalent where response to standard care was similar (14% CR, 21% PR) but 57% of those with a complete response would go on to alloSCT and 44% of those with a PR would receive alloSCT. However, it was noted that individual clinicians have small numbers of these patients so these percentages are estimates only.

The CS noted that clinicians surveyed were largely in agreement with the data in Cheah[7] and Lafferty[21] compared to clinical practice including a PFS of 3.5 months and OS of 25.2 months reported in Cheah. However, three clinicians suggested an OS of 12 months based on their practice. Furthermore, although the clinicians accepted the findings of this study, many reported no access to investigational agents which are included in Cheah et al.[7]

ERG comment:

- The company made efforts to apply the appraisal to a UK context with the use of the clinician survey. However due to the rarity of the disease at this stage clinicians did not see many patients per year (most commonly 3 or 4). Hence duration of treatment and percentage processing to alloSCT are estimates based on sparse data.

KEYNOTE-013

The company provided an overview of the safety results of a phase 1b (single arm) trial of pembrolizumab in patients with relapsed or refractory disease.[22] In this trial (KEYNOTE-013) patients had relapsed after, were considered ineligible for, or had refused autoSCT. All 31 patients had progressed on or after treatment with BV. The CS stated that the dosing of pembrolizumab does not support the EMA recommendation so it was excluded from the decision problem. In this trial, pembrolizumab was administered intravenously at a dose of 10 mg/kg every two weeks.

The company stated that AEs of any grade and attribution were reported in 30 0f 31 patients (97%). Overall, 68% of patients experienced one or more AEs that were deemed related to treatment. There were no grade 4 treatment-related AEs and no deaths related to study treatment. The publication

associated with this trial also provided further details on efficacy. The CR rate was 16% (90% CI, 7% to 31%). In addition, 48% of patients achieved a partial remission, for an overall response rate of 65% (90% CI, 48% to 79%). (70% of the responses lasted longer than 24 weeks (range, 0.14+ to 74+ weeks), with a median follow-up of 17 months. The progression-free survival rate was 69% at 24 weeks and 46% at 52 weeks.

ERG comment:

- As the dosing regimen of KEYNOTE-013 did not reflect the EMA recommendation for pembrolizumab, the company appropriately provided details of this trial as supplementary information only and did not use it to inform modelling.

4.2.4 Ongoing trials

KEYNOTE-087 is an ongoing trial but the company stated that all available data had been included in the submission. They further stated in response to clarification that [redacted].[10]

Two further trials were mentioned as being ongoing: KEYNOTE-204 and NCT03077828. KEYNOTE-204 is an ongoing, randomised, non-blinded study of pembrolizumab versus BV in patients with relapsed or refractory cHL. The company stated that KEYNOTE-204 was not within the indication/ license in the submission. NCT03077828 is a single arm, open-label phase II study of pembrolizumab together with a chemotherapy regimen "ICE" (ifosfamide, carboplatin, and etoposide) for the treatment of relapsed/refractory cHL. The company indicates that those patients in this trial who have received BV prior to enrolment may be relevant to the current submission. However the estimated study completion date is February 2020.[1]

ERG comment:

- The ERG believes that none of the ongoing studies could have informed the current submission.
- Further analysis of the KEYNOTE-087 trial may be informative particularly for assessing longer-term OS.

4.3 Critique of trials identified and included in the indirect comparison and/or multiple treatment comparison

The original search (CS, page 45) resulted in one relevant citation: the CSR for KEYNOTE-087. No comparator studies were found. The updated search (CS, page 46) found three more citations for the KEYNOTE-087 trial, but again, no comparator studies. Finally, a separate search for observational studies (CS, page 46) retrieved one relevant study: Cheah et al. 2016.[7]

Cheah et al. 2016 was also used as the comparator study in TA462 (ID972 - Nivolumab for treating relapsed or refractory classical Hodgkin lymphoma). In this appraisal the population of interest was patients who had had previous autologous stem cell transplant and brentuximab vedotin (i.e. cohort 1 in the current appraisal). In TA462, the committee concluded (FAD, point 4.7):

> "The committee considered whether the population and composition of treatments in the Cheah study reflected clinical practice in the UK. The committee noted that the study population partially matched the population of interest because around 70% of patients had previous autologous stem cell transplant and brentuximab vedotin. The committee noted a lack of detail on the precise combinations of chemotherapies given as standard of care in the study, and the inclusion of platinum-based therapies and 'other alkylators'. It considered that the study may not reflect UK practice, particularly regarding subsequent rates of allogeneic stem cell transplant. The committee noted that in response to

consultation, the company had explored UK standard-of-care data from the Haematological Malignancy Research Network and surveyed clinicians actively treating relapsed or refractory classical Hodgkin lymphoma in the UK. The committee considered that both the network data and the clinician survey somewhat supported the Cheah study as reflecting UK practice, but it recognised that the data were limited. The committee concluded that the Cheah study was the best available evidence for standard of care and considered it appropriate for its decision-making, but overall the clinical effectiveness of nivolumab compared with standard of care was highly uncertain because the comparator data may not fully represent UK clinical practice."[6]

ERG comments:

- This means the committee accepted that Cheah[7] is appropriate as a comparator study for people with relapsed or refractory classical Hodgkin lymphoma who have received autologous stem cell transplant and brentuximab vedotin, i.e. cohort 1 in the current appraisal.
- As mentioned above, in Cheah et al. 2016, 70% of patients had received previous autologous stem cell transplant and brentuximab vedotin. In Table 1 of their publication, Cheah et al. 2016[7] report baseline characteristics of 97 included patients before commencement of brentuximab vedotin. Of these 97 patients, 70 had previous stem cell transplantation (SCT), 66 had autoSCT and 4 had alloSCT. The remaining 27 patients did not undergo consolidative transplant; for these, the primary reason was failure to respond to therapy (n = 21, 75%), age or co-morbidities (n = 1, 4%), failed mobilization (n = 1, 4%), patient decision (n = 1, 4%), financial reasons (n = 1, 4%) or reason unknown (n=2, 7%).[7] The CS reports ITT data from Cheah et al., i.e. data for the whole population, with and without transplant. We asked the company to provide separate cohort analyses, using separate data from the two cohorts (with and without transplant) in the Cheah et al. study (Clarification Letter, question A16). The company responded that "MSD do not have access to individual patient level data for Cheah et al. and therefore it was not possible to determine cohorts using the same inclusion criteria as were applied to cohorts 1 and 2 in KEYNOTE-087".[10]
- Using the full Cheah et al. 2016 population as a comparator for cohort 2 is problematic. First of all, only 27 out of 97 patients (28%) did not undergo consolidative transplant. In addition, as shown in Table 4.11 there are differences between the population in cohort 2 and Cheah regarding age, ECOG score, Baseline B symptoms, Haemoglobin, Lymphocytes, Albumin, White cell count and Bulky Lymphadenopathy.
- In summary, the Cheah population is a mixture of both cohort 1 and 2 (as defined in the scope) and is probably most comparable to cohort 1 in the KEYNOTE-087 trial. Because KEYNOTE-087 shows that results are more favourable in cohort 1 compared to cohort 2, using the total population from Cheah as the comparator means results of the naïve comparison will probably overestimate the effect of pembrolizumab in cohort 1 and underestimate the effect of pembrolizumab in cohort 2.

Table 4.11: Baseline characteristics of patients in the included studies

Characteristic		KEYNOTE-087, Cohort 1	KEYNOTE-087, Cohort 2	Cheah et al. (2016)
Treatment		Pembrolizumab 200mg		Mix of therapies including chemotherapy, and investigational agents
Number of patients		69	81	97[α] or 89[β]
Age (median)		34.0	40.0	32[β]
Age >45 (%)		25%[γ]	42%[γ]	14 (14%)[β]
Female (%)		33 (47.8%)	38 (46.9%)	46 (47%)[α]
ECOG	0	29 (42.0%)	44 (54.3%)	33 (41%)[β]
	1	39 (56.5%)	37 (45.7%)	44 (54%)
	2	1 (1.4%)	0 (0.0%)	3 (4%)
Stage	1	NR	NR	2 (3%)[β]
	2	NR	NR	25 (30%)
	3	NR	NR	18 (21%)
	4	NR	NR	39 (46%)
Baseline B symptoms		22 (31.9%)	26 (32.1%)	7 (8%)[β]
Haemoglobin <105 g/l		35%[γ]	27%[γ]	18 of 51 (35%)[β]
Lymphocytes <0.6 × 10^9/l		19%[γ]	15%[γ]	19 of 46 (41%)[β]
White cell count >15 × 10^9/l		9%[γ]	17%[γ]	4 of 82 (5%)[β]
Albumin <40 g/l		48%[γ]	49%[γ]	23 of 82 (28%)[β]
Any extranodal site		56%[γ]	41%[γ]	31 of 88 (35%)[β]
Maximum tumour diameter ≥4 cm		49%[γ]	42%[γ]	18 of 69 (26%)
Bulky Lymphadenopathy		5 (7.2%)	11 (13.6%)	15 (37%)[α]
Bone marrow involvement		3 (4.3%)	5 (6.2%)	NR
Disease status - relapse		46 (66.7%)	24 (29.6%)	NR
Disease status – refractory		23 (33.3%)	57 (70.4%)	NR
Previous BV therapy		69 (100.0%)	81 (100.0%)	89 (100%)[β]
Prior autoSCT		69 (100.0%)	0 (0.0%)	66 of 97 (68%)[α]
Prior radiation		31 (44.9%)	21 (25.9%)	NR
Median no. of prior line of therapy		4	4	3[α]

Source: CS, Table 53, page 137-139[1], CS Appendix 8, and Tables 1 and 2 in Cheah et al. 2016[7].
[α] Sample before commencement of BV (Table 1); [β] Sample at the time of documented progression following therapy with BV (Table 2) – not all characteristics were available from the same sample; [γ] From Appendix 8.
BV = Brentuximab vedotin; ECOG = Eastern Cooperative Oncology Group; N/A = Not applicable; NR = Not reported; SCT = Stem cell transplant

Another issue regarding the population in the Cheah et al. 2016 study is that patients received a wide variety of treatments (see Table 4.12).

Cheah et al. (2016)[7] conducted a retrospective review of their institutional database (at the MD Anderson Cancer Center, Texas) to identify patients who had been treated with BV between June 2007 and January 2015. To be included in the study patients had to meet the following criteria:

- A histologically confirmed diagnosis of classical Hodgkin lymphoma
- Treatment with BV for relapsed Hodgkin lymphoma
- Disease progression at any time after treatment with BV

The aim of the study was to determine PFS and OS following disease relapse after BV therapy. Secondary outcomes were to analyse the efficacy of subsequent therapeutic strategies and to explore candidate prognostic factors for PFS and OS.

Cheah et al. (2016) report that 66/97 (68%) had prior ASCT and 4 (4%) had prior alloSCT conducted at the time of second remission. Data were available on subsequent therapy for 83 patients with disease progression following BV therapy and these data are reproduced below in Table 4.12. The proportion of patients who had prior ASCT among the 83 patients with disease progression is not reported.

Table 4.12: Therapies received by patients in the Cheah et al. study who had disease progression following BV therapy

Treatment	n	Evaluated	CR (%)	PR (%)	ORR (%)	mPFS	mOS
Investigational agent	28	28	4 (14)	3 (11)	7 (25)	2.4 (months)	47.7 (months)
Gemcitabine	15	12	4 (27)	4 (27)	8 (53)	2.1	NR
Bendamustine	12	11	2 (17)	4 (33)	6 (50)	3.7	34.0
Other alkylator	6	4	1 (17)	1 (17)	2 (33)	5.0	9.5
BV retreatment	6	4	0 (0)	2 (33)	2 (33)	3.5	10.4
Platinum based	4	4	0 (0)	1 (25)	1 (25)	0.9	25.2
ASCT	3	3	1 (33)	0 (0)	1 (33)	-	11.9
Other	5	1	0 (0)	0 (0)	0 (0)	-	24.9
Overall	79	67 (85%)	12 (15)	15 (16)	27 (34)	3.5	25.2
No treatment received	4 due to poor performance status and/or patient decision						
TOTAL	83						

ASCT = autologous stem cell transplant; BV = brentuximab vedotin; CR = complete response; mOS = median overall survival; mPFS = median progression-free survival; ORR = objective response rate; PR = partial response

In TA462,[6] the company performed two analyses: using the overall Cheah population (i.e. including efficacy from all the treatments listed above) and using the Cheah population but excluding efficacy data for the n=28 patients who received investigational agents. This was because the 'Investigational Agent' group could have included nivolumab. According to the ERG report for TA462 "only a couple of patients in the study received PD-1 inhibitors (although numerical data to support this statement were not provided)."[11] Results of these analyses showed that excluding data for patients who received investigational agents, improved effectiveness results for Cheah. Results for pembrolizumab for the current STA excluding investigational agents are presented below in Section 4.4.2.

4.4 Critique of the indirect comparison and/or multiple treatment comparison

4.4.1 Methodology of the indirect comparison

The company presents indirect comparisons for the following outcomes: response rate ORR (CR+PR) and survival (PFS).

The company states that it was not possible to consider OS within the long-term model structure in those who do not receive an alloSCT due to a lack of events during the follow-up period. In addition, no formal method of data analysis was proposed by the company for AEs or HRQoL, as these data are not available within the comparator study Cheah et al. (2016).

For each outcome, the company performed two types of analyses: a naïve indirect comparison (IC) and a matched adjusted indirect treatment comparison (MAIC).

A naïve IC was used to compare pembrolizumab using data from KEYNOTE-087 with standard of care (SoC) using data from Cheah et al. (2016). This was a comparison of two single arms, due to the lack of a randomised comparison. PFS was compared using a Cox proportional hazards (PH) model to obtain a naïve unadjusted hazard ratio (HR) for two scenarios:

1. From study initiation to most recent observation
2. From study initiation to week 12

ORR was compared between pembrolizumab and SoC using a chi-squared test for the same time periods as the PFS analysis.

The MAIC used weighting to match the IPD from KEYNOTE-087 to the summary data from Cheah et al. (2016). The methods provided in NICE DSU report 18 (Methods for population-adjusted indirect comparisons in submissions to NICE)[8] were used and the weights applied to the KEYNOTE-087 data were derived from the inverse odds of being in pembrolizumab compared to SoC.

The initial matching used all variables for which data were available in both KEYNOTE-087 and Cheah et al. (2016). In cases where the algorithm used to estimate the weights did not converge using the full set of baseline characteristics, variables were removed in stepwise fashion in a predetermined order until convergence was achieved.

Weights from the propensity model were then applied to a Cox regression model with the same structure as used for the naïve IC to obtain population-adjusted HRs for the same two scenarios:

1. From study initiation to most recent observation;
2. From study initiation to week 12.

For ORR, the same method was used to estimate weights for each separate comparison. Weighted contingency tables and chi-squared test for difference between pembrolizumab and SoC were used to estimate odds ratios.

The company states they conducted a feasibility assessment that focused on two areas: the compatibility of included studies and the data published on potential confounders i.e. the extent to which adjustment could be made to ensure exchangeability.[23, 24] Compatibility was assessed by comparing study design characteristics such as inclusion and exclusion criteria, study endpoints and methods for outcomes assessments (CS, Section 4.10.12). However, the results of this assessment were not fully reported apart from in tables summarising baseline characteristics for each study. The compatibility assessment by the ERG is as follows:

- *Study design characteristics such as inclusion and exclusion criteria*

Cheah et al. (2016)[7] was a retrospective study including patients with (i) a histologically confirmed diagnosis of cHL, (ii) treatment with BV for relapsed HL and (iii) subsequent disease progression at any time after treatment with BV. Patients' treated with BV as part of frontline HL therapy was excluded.

KEYNOTE-087 was a prospective study including patients with relapsed (disease progression after most recent therapy) or refractory (failure to achieve CR or PR to most recent therapy) de novo cHL and

(1) Have failed to achieve a response or progressed after autoSCT. Patients must have relapsed after treatment with or failed to respond to BV post autoSCT (cohort 1); or

(2) Were unable to achieve a CR or a PR to salvage chemotherapy and did not receive autoSCT. Patients must have relapsed after treatment with or failed to respond to BV (cohort 2).

As can be seen from Table 4.11, there are differences in baseline characteristics between the KEYNOTE-087 cohorts and the Cheah et al. (2016)[7] study, regarding age, ECOG score, baseline B symptoms, Lymphocytes, White cell count, Albumin level, extranodal site, tumour diameter, and Bulky Lymphadenopathy.

- *Study endpoints*
  Cheah et al. (2016)[7] reports PFS, OS, and ORR, CR and PR. OS and PFS were reported as median survival times in months and the CR rate, PR rate and ORR as percentages. The same outcomes are reported in the CS for pembrolizumab. However, OS was not included in the indirect comparisons due to due to a lack of events during the follow-up period.

- *Methods for outcomes assessments*
  The ERG notes that there were differences in how PFS was defined between the pembrolizumab study and Cheah et al. (2016)[7]. In the pembrolizumab study (KEYNOTE-087), PFS was defined as the time from first treatment to disease progression, as assessed by BICR per IWG response criteria for malignant lymphoma and by site review or death due to any cause, whichever occurred first (CSR, page 3). In contrast, the PFS definition in Cheah et al. (2016)[7] was the time in months measured from date of confirmed disease relapse following BV to disease progression or death.

  Regarding treatment response, there are also differences in definitions. Cheah et al. (2016) only state that 'treatment responses were determined according to the 2014 Lugano Classification.[25]'. In the pembrolizumab study (KEYNOTE-087), best Overall Response Rate (ORR) was defined as the proportion of patients in the analysis population who have complete remission (CR) or partial remission (PR) using IWG criteria (Cheson 2007[26]) at any time during the study. In KEYNOTE-087 response at 12 weeks follow-up was also assessed.

In summary, the compatibility assessment shows that there are some differences between the two studies regarding baseline characteristics and methods of outcomes assessment. However, the differences in baseline characteristics are more of a concern for the results of the naïve comparison as this is a comparison of two different studies. The MAIC is less affected as the two studies have been matched as part of the analysis method to try and make them comparable at baseline. Differences in methods of outcome assessment are due to the two different study designs (single-arm prospective study vs. retrospective study) and these are a concern as the individual patient data were not available for Cheah et al. (2016) so the OS and PFS outcomes could not be recalculated to match the methods used in KEYNOTE-087.

The second part of the feasibility assessment conducted by the company focused on the data published on potential confounders i.e. the extent to which adjustment could be made to ensure exchangeability.[23, 24] The baseline characteristics from all patients with available data at the time of documented progression following treatment with BV were applied and used for matching (the number of patients with data varied from 89 for age to 46 for lymphocyte count). According to the company, matching was conducted using all variables for which data were available in both KEYNOTE-087 and Cheah et al. (2016). Appendix 8 of the CS lists the variables included in the matching exercise, these are: ECOG >0 (%), B symptoms (%), Age >45 (%), Albumin <40 g/l (%), Haemoglobin <105 g/l (%), Lymphocytes $<0.6 \times 10^9$ (%), White blood cells $>15 \times 10^9$ (%), Max Tumour Diameter >4 cm (%), Any extranodal site (%), Female (%), and Prior lines (mean/median). As can be seen in Table 4.11, apart from Bulky Lymphadenopathy and prior autoSCT, these are indeed the only available variables for adjustment. The company does not explain why Bulky Lymphadenopathy and prior autoSCT were not included in the matching but did state that in cases where the weighting algorithm did not converge then variables were removed in a stepwise fashion in a predetermined order until convergence was achieved.

The naïve IC results should be treated with caution due to the differences in patient populations and study design between KEYNOTE-087 and Cheah et al. (2016) and the fact that they are a comparison of single-arms and not based on randomised trials. There was no attempt to match the populations used in this analysis so the results are from comparisons of different treatment groups from two single-arm studies of different designs (one prospective and one retrospective).

The MAIC used recommended methods and appears to have been conducted correctly. Initially all baseline variables which were available in both studies were included in the matching algorithm. Variables were only excluded from the matching if there were problems with model convergence. Most analyses only excluded one variable 'median prior lines', but the analysis of ORR for cohort 1 in the 12-week scenario only included four variables in the matching model. The reason for this reduced model was not provided. The baseline characteristics pre- and post-matching for each study and outcome are presented in Appendix 8 and show that a satisfactory match was obtained between KEYNOTE-087 and Cheah et al. (2016) for all eligible variables.

The ERG could not reproduce the MAIC for checking as only the IPD for KEYNOTE-087 were provided by the company. The data for Cheah et al. (2016) were not provided even though it was used in the analysis and the ERG had requested all data and the corresponding R code in the clarification letter.

According to DSU report 18 (Methods for population-adjusted indirect comparisons in submissions to nice)[8] "companies deploying MAIC or STC are not only arguing that the treatment effect is dependent on the population, but they are further assuming that the target population is closer to that represented in the comparator trial than in their own trial." In this case, this means that the MAIC analysis is based on the population characteristics as in the Cheah et al. (2016) study. As stated above, in TA462, the committee "considered that the study may not reflect UK practice, particularly regarding subsequent rates of allogeneic stem cell transplant" and "The committee concluded that the Cheah study was the best available evidence for standard of care and considered it appropriate for its decision-making, but overall the clinical effectiveness of nivolumab compared with standard of care was highly uncertain because the comparator data may not fully represent UK clinical practice."[6].

According to DSU report 18,[8] unanchored indirect comparisons (i.e. those based on single-arm studies) are susceptible to large amounts of systematic error unless all prognostic variables and effect modifiers are accounted for in the propensity score weighting model. However, in the current MAIC the company was dependent on the variables reported in Cheah et al. (2016) and these are unlikely to be all relevant

prognostic variables and effect modifiers. In addition, some variables had to be dropped from some models to enable the models to converge. DSU report 18 recommended that information should be provided on the level of bias likely to be introduced as a result of any covariates that are unaccounted for. However, the company did not provide this due to a "lack of studies in the patient population relevant to this analysis". They did not comment on the degree of systematic error within the MAIC estimates. Therefore, the results are likely to contain systematic error but it is not possible to estimate the size of the potential error.

Summary regarding indirect comparison with Cheah et al. 2016[7]:

- There are problems with compatibility of the two studies (KEYNOTE-087 and Cheah et al (2016)) regarding baseline characteristics and methods of outcomes assessment, although this has a greater impact on the results of the naïve IC as the MAIC does try to match the two groups prior to analysis.
- Using the full Cheah et al. (2016) population as a comparator for cohort 1 is probably acceptable given the committee's discussion in TA462.
- Using the full Cheah et al. 2016 population as a comparator for cohort 2 is problematic, because only 27 out of 97 patients (28%) did not undergo consolidative transplant and there are differences between the population in cohort 2 and Cheah regarding age, ECOG score, Baseline B symptoms, Haemoglobin, Lymphocytes, Albumin, White cell count and Bulky Lymphadenopathy (see Table 4.10).
- The MAIC analysis is based on the population characteristics as in the Cheah et al. (2016) study. These characteristics may not fully represent UK clinical practice.
- The naïve IC results are from two different patient populations and study designs and are likely to be biased as they are not based on data from RCTs.
- The results of the MAIC are likely to include systematic error and the relative treatment effects are only estimated for the target population in the comparator trial (Cheah et al. (2016)).
- Both the naïve IC and MAIC have major limitations and neither are fully reliable for decision making. In the company model and in the ERG analysis the naïve IC is used in the base case and the MAIC in sensitivity analyses.

4.4.2 Results of the indirect comparison

Results for PFS and ORR are reported in Tables 4.13 and 4.14, respectively. The company also presented results for CR and PR (CS, Tables 31-34, pages 98-100).

Almost all results for PFS show a significant benefit for pembrolizumab versus SoC. The only exception is the naïve comparison in cohort 1 in the 12-week scenario, this shows a non-significant difference favouring pembrolizumab but the upper 95% confidence limit only just crosses one. Likewise, all results for ORR significantly favour pembrolizumab over SoC. Results of the naïve comparison are similar to MAIC.

This analysis excluded baseline data from median prior lines in the matching.

Table 4.13: Summary of comparisons of progression-free survival for pembrolizumab versus SoC

Cohort	Comparison	Sample size, n	Pembrolizumab		Hazard ratio (95% CI)
			Events, n	Censored, n	
Entire study scenario					
1	Naïve	[redacted]	[redacted]	[redacted]	[redacted]
	MAIC		[redacted]	[redacted]	[redacted]
2	Naïve	[redacted]	[redacted]	[redacted]	[redacted]

	MAIC		[redacted]	[redacted]	[redacted]
12-Week scenario					
1	Naïve	[redacted]	[redacted]	[redacted]	[redacted]
	MAIC		[redacted]	[redacted]	[redacted]
2	Naïve	[redacted]	[redacted]	[redacted]	[redacted]
	MAIC		[redacted]	[redacted]	[redacted]
Source: CS, Tables 27 and 28, page 96. CI = confidence interval; SoC = standard of care					

Table 4.14: Summary of comparisons of objective response rates for pembrolizumab versus SoC

Cohort	Comparison	Sample size, n	ORR with pem	ORR with SOC	Odds ratio (95% CI)
Best overall response					
1	Naïve	[redacted]	[redacted]	[redacted]	[redacted]
	MAIC	[redacted]	[redacted]	[redacted]	[redacted]
2	Naïve	[redacted]	[redacted]	[redacted]	[redacted]
	MAIC	[redacted]	[redacted]	[redacted]	[redacted]
12 Weeks					
1	Naïve	[redacted]	[redacted]	[redacted]	[redacted]
	MAIC	[redacted]	[redacted]	[redacted]	[redacted]
2	Naïve	[redacted]	[redacted]	[redacted]	[redacted]
	MAIC	[redacted]	[redacted]	[redacted]	[redacted]
Source: CS, Tables 29 and 30, page 97. CI = confidence interval; ORR = objective response rate; pem = pembrolizumab; SOC = standard of care					

In the clarification letter we asked the company to perform an analysis using the Cheah population but excluding efficacy data for the n=28 patients who received investigational agents (as in TA462) (Clarification question A13). The company was not able to provide such an analysis for PFS, but was able to provide this analyses for response (ORR, CR and PR).

Results for ORR are presented in Table 4.15 below, these results still show a significant benefit for pembrolizumab versus SoC although on the whole less favourable. The analysis of the entire study period excluded baseline data from median prior lines in the matching. For the analysis up to 12 weeks the results for cohort 1 only included four variables in the model (ECOG, B symptoms, age and albumin) but all variables except median prior lines, were included in the model for cohort 2.

Table 4.15: Summary of comparisons of objective response rates for pembrolizumab versus SoC after removing investigational agents

Cohort	Comparison	Sample size, n	ORR with pem	ORR with SOC	Odds ratio (95% CI)
Best overall response					
1	Naïve	[redacted]	[redacted]	[redacted]	[redacted]
	MAIC	[redacted]	[redacted]	[redacted]	[redacted]
2	Naïve	[redacted]	[redacted]	[redacted]	[redacted]
	MAIC	[redacted]	[redacted]	[redacted]	[redacted]
12 Weeks					
1	Naïve	[redacted]	[redacted]	[redacted]	[redacted]
	MAIC	[redacted]	[redacted]	[redacted]	[redacted]
2	Naïve	[redacted]	[redacted]	[redacted]	[redacted]
	MAIC	[redacted]	[redacted]	[redacted]	[redacted]
Source: Response to clarification, Question A13, Tables 1 and 2. CI = confidence interval; ORR = objective response rate; pem = pembrolizumab; SOC = standard of care					

4.5 Additional work on clinical effectiveness undertaken by the ERG

No further additional work was undertaken by the ERG.

4.6 Conclusions of the clinical effectiveness section

The CS includes a systematic review of the evidence for pembrolizumab and its comparators in patients with classical Hodgkin Lymphoma who have either received autoSCT and BV or BV alone due to autoSCT being unsuitable. No relevant randomised trials were identified.

One study of the efficacy and safety of pembrolizumab was identified (KEYNOTE-087) and this formed the basis of the submission. KEYNOTE-087 is a well-conducted single arm trial including 150 patients relevant to this appraisal. This ongoing multicentre trial includes three UK centres (14 UK patients). The main trial in the CS covers both cohorts of interest (cohort 1: people with relapsed or refractory cHL who have received autologous stem cell transplant and brentuximab vedotin and cohort 2: patients who have received brentuximab vedotin when autologous stem cell transplant is not a treatment option). The company presented data based on a median follow up of 15.9 months. The median time on treatment was [redacted] days for cohort 1 and [redacted] days for cohort 2.

The primary outcome was overall response rate (ORR) as assessed by independent committee. ORR was [redacted] in cohort 1 and [redacted] in cohort 2. In cohort 1 [redacted] of patients had a complete response and in cohort 2 this figure was [redacted] Median progression free survival (PFS) in cohort 1 was [redacted]. In cohort 2 it was [redacted]. Median overall survival (OS) was [redacted]. At 12 months survival was [redacted] in cohort 1 and [redacted] in cohort 2. In cohort 1 [redacted] of patients had one or more adverse events. In Cohort 2 [redacted] of patients had one or more adverse events. The company noted that most AEs were low grade ([redacted] and [redacted] grades 3 to 5 in cohort 1 and 2 respectively). In cohort 1 [redacted] of AEs were classed as serious and in cohort 2 [redacted]. The most common adverse events were pyrexia, cough, fatigue, diarrhoea and vomiting. The company conducted post-hoc analyses of response. The main difference between this post-hoc analysis and the primary analysis of response, referred to as 'best' response rate, is that response was determined at a single time point for each patient i.e. 12 weeks as opposed to any time point up to the point of progression. Data in the form of the proportion who respond by a specific time point was required in order to apportion patients that were progression-free into the cost effectiveness model. The ERG noted that overall response rates were lower at 12 weeks than over the course of the trial ([redacted]

The most important methodological aspect to note are that although the trial was well conducted, it represents a low level of evidence. It is a phase II, single arm, non-comparative study which by its design has serious limitations. We cannot know whether the outcomes observed are a true reflection of the intervention as the role of natural history and baseline characteristics is not taken into account. As treatment is known to participants, clinicians and assessors this can lead to bias in the delivery of the intervention and the reporting of outcomes. Other limitations in applying the results of the trial to UK practice include:

- Although the study had an adequate follow-up (median 15.9 months) for the primary outcome (ORR), median progression free survival was immature and median overall survival [redacted].
- The trial has only 150 relevant participants so the evidence base for this appraisal is small. Patients over 65 are not well represented. Furthermore, a small number of patients were from the UK (14) so the trial may not totally reflect the UK population and setting. It is recognised,

however, that the population matching the scope of this appraisal from which to draw participants is in itself small.

- In clinical practice, for those who are suitable, pembrolizumab represents a bridge to alloSCT, a potentially curative treatment. However the company submission stated that 'KEYNOTE-087 was not designed as a 'bridging' study, therefore the uptake of alloSCT was very low overall across cohorts 1 and 2.'[1] The company further stated that 'the use of stem cell transplant would have been at the discretion of the treating physician on a per patient basis.'[10]

As KEYNOTE-087 did not have a comparator group the company identified a comparative observational study from the literature (Cheah et al 2016[7]). This is a retrospective USA database study in which patients received the following types of therapy: investigational agent(s), gemcitabine, bendamustine, any other alkylator, BV retreatment, platinum based treatment, autoSCT or alloSCT, or other treatment. The company has not provided separate data for comparators; instead a combined data set has been provided for multiple comparators, some of which are within the scope and others not. In the previous appraisal of nivolumab (TA462)[6], the committee concluded that "the Cheah study was the best available evidence for standard of care and considered it appropriate for its decision-making, but overall the clinical effectiveness of nivolumab compared with standard of care was highly uncertain because the comparator data may not fully represent UK clinical practice." However, the ERG is not aware of a more appropriate source of data for the comparator population for this appraisal.

The company performed two types of analyses: a naïve indirect comparison between KEYNOTE-087 and Cheah and a matched adjusted indirect treatment comparison (MAIC) of the two studies. The ERG identified problems with compatibility of the two studies regarding baseline characteristics and methods of outcomes assessment. In the MAIC the company adjusted for potential confounding variables so that the KEYNOTE-087 study more closely resembled the Cheah study.

Almost all results for PFS show a significant benefit for pembrolizumab versus SoC. The only exception is the naïve comparison in cohort 1 in the 12-week scenario, this shows a non-significant difference favouring pembrolizumab but the upper 95% confidence limit only just crosses one. Likewise, all results for ORR significantly favour pembrolizumab over SoC. Results of the naïve comparison are similar to MAIC. However, the results of the naïve comparison and MAIC are not reliable because they are likely to contain systematic error but it is not possible to estimate the size of the potential error. Both the naïve IC and MAIC have major limitations when used for decision making.

Although not listed as a comparator in the NICE scope and not referenced in the CS, nivolumab has recently received approval from NICE for this condition. It is recommended 'as an option for treating relapsed or refractory classical Hodgkin lymphoma in adults after autologous stem cell transplant and treatment with brentuximab vedotin.'[6] This represents cohort 1 of this appraisal. It will be important to compare the efficacy and safety of nivolumab and pembrolizumab for this cohort.

5. COST EFFECTIVENESS

5.1 ERG comment on company's review of cost effectiveness evidence

This section refers to the review of cost effectiveness analysis studies comparing pembrolizumab to comparator therapies in the treatment of RRcHL, as well as the review of studies on health-related quality of life and resource requirements and costs associated with treatment of the patient population, as presented in the CS chapter 5.1[1] and Appendix 12 of the CS.[27]

5.1.1 Objective of cost effectiveness review

All searches presented in the CS relating to cost effectiveness will be summarised and commented on in the following paragraphs.

Objective of cost effectiveness analysis search and review

Three SLRs were performed by the company with the aim of identifying all literature supporting the development and population of a model of patients with relapsed or refractory classic Hodgkin Lymphoma, treated with pembrolizumab. Within the SLRs, the company executed a single set of searches to address the following areas: (1) cost-effectiveness studies of comparator therapies vs. pembrolizumab, (2) health-related quality of life (HRQoL) in the patient population, and (3) resource requirements and costs associated with treatment.

The CS reported that searches were carried out in July 2017. Searches were limited to studies published from 2001-2017, but were not limited by language. Searches were carried out on the following databases: Embase, MEDLINE, MEDLINE in-Process (searched via Pubmed), HTA and NHS EED via the Cochrane library and EconLit. Searches contained facets to identify relevant studies regarding the costs, HRQoL and resource use identification of classical Hodgkin Lymphoma. Searches were carried out in line with the NICE 2013 guide to the methods of technology appraisal Sections 5.2.2 and 5.2.4.[28] Supplementary searches of the following conference proceedings were reported for the previous two years: American Society of Clinical Oncology (ASCO), the European Society for Medical Oncology (ESMO) and the International Society for Pharmacoeconomics and Outcomes Research (ISPOR). The CS also reported that the NICE website was searched in order to identify relevant information from previous submissions not otherwise captured.

ERG comment: The ERG comments are in relation to (a) well reported and reproducible searches and (b) limitation around simultaneous search of two databases.

(a) The majority of searches in Appendix 12 were well reported and easily reproducible. In the original submission, strategies for EconLit, the ASCO, ESMO and ISPOR conference proceedings and a search of the NICE website were not included in Appendix 12, these were provided by the company in their response to clarification.[10]

(b) The ERG asked the company to clarify whether the MEDLINE/Embase strategy reported in Appendix 12, was a single search conducted simultaneously over both the Embase and MEDLINE individual databases or a single search of Embase conducted on the understanding that it now contains all records from MEDLINE. The company responded that "The first search strategy covers evidence from both Embase and MEDLINE using the embase.com interface".[10] The ERG took this as confirmation that a simultaneous search of the two databases had taken place. This approach has limitations when using subject heading terms. It appeared that only Embase subject heading terms (Emtree) were used in the search strategy, and although simultaneous searching of embase.com should automatically identify and search for equivalent MEDLINE subject heading terms (MeSH), as

the ERG does not have access to Embase.com for testing it is not clear if this is the case for all potentially useful MeSH terms. Given the potential limitations of this approach, the ERG considered it preferable to search each database separately, or at least to ensure inclusion of both Emtree and MeSH terms in the search strategy. However, given the additional searches, this is unlikely to have affected the overall recall of results.

5.1.2 Inclusion/exclusion criteria used in the study selection

Complete lists of inclusion and exclusion criteria are provided in the CS (CS Table 52)[1] and in CS Appendices 13 and 15.[27] Below a summary of the inclusion criteria is provided:

Population: adult patients with relapsed/refractory cHL, irrespective of age or gender (mixed populations were excluded unless subgroup data on the population of interest was provided).

Intervention and comparator: No restriction, all pharmacological interventions to be captured.

Outcomes:

1) Studies including a comparison of benefits and costs between the intervention and comparator arms. Results expressed in incremental costs, incremental cost effectiveness ratio (ICER), quality-adjusted life-years (QALYs), life-years gained (LYG) or other measures of effectiveness additional to costs.
2) Studies reporting health state utilities of interest
3) Studies reporting costs

Study type:

1) Cost effectiveness analysis, cost utility analysis, cost benefit analysis, cost minimisation analysis, budget impact analysis, cost consequence analysis,
2) Studies using European Quality of Life-5 Dimensions (EQ-5D), European Organisation for Research and Treatment Cancer Quality of Life Questionnaire (EORTC QLQ-C30), short form 36 health survey (SF36), health utility index (HUI), Visual Analogue Scale (VAS), Time trade off (TTO) or Standard Gamble (SG),
3) Cost studies, surveys, burden of disease and resource use studies

Other: studies published from 2001 onwards, full text in English language and reporting UK specific data (cost data from other countries allowed).

5.1.3 Included/excluded studies in the cost effectiveness review

A total of 2,051 references were identified in the SLRs.

(1) Of 848 identified cost effectiveness references, 52 duplicates were removed and 796 abstracts were screened which led to the exclusion of 694 articles. Consequently, 102 full texts were screened, all of which had to be excluded (see CS p. 135 for the PRISMA diagram[1]). No cost effectiveness studies were included through other searches.

(2) Of 1,236 references identified on HRQoL and utilities, 95 duplicates were removed and 1,141 abstracts were screened which led to the exclusion of 1,091 articles. Subsequently, 50 full-texts were screened of which 46 were excluded and two studies from four publications were included (see CS p. 187 for the PRISMA diagram[1]).

(3) Of 882 identified cost articles, 52 duplicates were removed and 830 abstracts were screened which led to the exclusion of 728 articles and the full-text screening of 102 articles. After the exclusion of 86

more articles and the inclusion of one article through conference searching, a total of 14 studies from 17 publications were included, one of them reporting UK-specific costs and resource use (see CS p. 187 for the PRISMA diagram[1] and Appendix 14 of the CS[27] for a list of studies included).

5.1.4 Conclusions of the cost effectiveness review

No cost effectiveness studies in patients with RRcHL were identified that met the inclusion criteria, therefore the company conducted a de novo health economic analysis. The majority of relevant utility studies identified did not use EQ-5D data and were thus inconsistent with the NICE reference case, or reported utilities not stratified by response. Disutilities of grade 3+ adverse events (AEs) were sourced from previous TAs (see Table 78 of the CS[1]). Fourteen cost studies were found to meet the inclusion criteria. As the updated publication of one of the identified cost studies (Radford 2017,[29]) was the preferred source of cost data in TA462[6] (Nivolumab for treating relapsed or refractory classical Hodgkin lymphoma), this reference was selected to inform the economic analysis. AE costs were computed from a weighted average of Healthcare Resource Group (HRG) code prices.

ERG comments:

The ERG agrees that in the absence of cost effectiveness studies performed on the population and intervention of interest from the literature, a de novo approach was necessary. It was, however, unclear why the company did not provide a complete overview of the publications included and excluded from their cost effectiveness, cost and resource and utility and HRQoL SLRs. Furthermore, the number of references found on EconLit was reported inconsistently in CS Appendix 12 [27] and PRISMA diagrams (CS pages 187 and 198). In their response to clarification question B2, the company explained that the PRISMA diagrams contain the correct number of publications. The ERG wishes to point out that the company prioritised aligning their sources with TA462 over using the results of their SLRs.

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

Table 5.1: Summary of the company's economic evaluation (with signposts to CS)

	Approach	Source/Justification	Signpost (location in CS)
Model	A state transition model with a decision tree element to predict response and alloSCT uptake	To provide an estimate of the lifetime costs and effectiveness of a "bridging" treatment to alloSCT or continued treatment with pembrolizumab.	Chapter 5.2.2
States and events	Health states week 0-12 (short-term component): - Progression free, consisting of complete response, partial response and stable disease - Progressed disease - Death Health states after 12 weeks (long-term component): - Alive post-alloSCT	The short-term health states capture initial treatment response determining alloSCT uptake. The long-term health states describe progression free survival, overall survival and response conditional on alloSCT uptake or continued use of pembrolizumab or SoC. A post-alloSCT progressed disease health state is missing with the justification that the	Chapter 5.2.2

	Approach	Source/Justification	Signpost (location in CS)
	- Progression free non-alloSCT - Progressed disease non-alloSCT - Death	implications of progression post-alloSCT are included in the post-alloSCT alive health state utilities and costs.	
Comparators	- SoC, consisting of chemotherapies (38.5%), treatment with investigational agents (43.1%) and bendamustine (18.5%) - BSC (only in scenario analysis)	SoC was included as a comparator as it contained combination chemotherapy such as gemcitabine, vinblastine and cisplatin, as defined in the scope [4]. Although also identified as a comparator in the scope [4], BSC was only used as a comparator in a scenario analysis. According to expert opinion, the use of BSC in UK practice is minimal.	Chapter 5.2.4
Population	Adult patients with RRcHL who have failed autoSCT and BV (cohort 1), or who are autoSCT ineligible and have failed BV (cohort 2).	This is consistent with the final scope issued by NICE and the population of the KEYNOTE-087 trial	Chapter 5.2.1
Treatment effectiveness	Due to the characteristic of pembrolizumab as a "bridging" treatment to alloSCT, treatment effectiveness was driven by the proportion of patients responsive to treatment at 12 weeks, allowing for alloSCT uptake. Pre-12 week effectiveness was informed by OS and PFS curves. Post-12 weeks, in the non-alloSCT pathway, OS and PFS curves were fitted. Post-12 weeks in the alloSCT pathway, OS was independent of prior treatment.	Proportional hazards were assumed to hold for all estimates. Comparative effectiveness and response were estimated by a naïve comparison of single-arm studies Cheah et al [7] and KEYNOTE-087. A matched indirect comparison was performed in a scenario analysis in order to avoid data loss in the base-case. Probabilities for alloSCT conditional on response states were elicited from UK clinical experts via two clinician surveys because the KEYNOTE-087 study was deemed non-generalizable to the UK setting. Post-alloSCT OS estimates were derived from Lafferty et al [21], because there was insufficient long-term data in the KEYNOTE-087 study. In the non-alloSCT post-progression	Chapter 5.3 Chapter 5.3.1

	Approach	Source/Justification	Signpost (location in CS)
		health state, mortality was based on Cheah et al.[7]	
Adverse events	Resource use, costs and utility decrement (one-off) were considered for AEs grade 3+	AEs with an incidence of >0% in either treatment arm, in line with TA462[6] were selected. Disutilities stemmed from literature sources used in previous TAs.[30, 31, 32, 33, 34, 35, 36]	Chapter 5.3.5
Health related QoL	Utilities are based on the KEYNOTE-087 study (using 12 week EQ-5D data only) in combination with (treatment specific) response rates. Moreover, utility values obtained from the literature were used.	Response-specific values elicited consistently with committee preference in TA462.[6]	Chapter 5.4.7 Chapter 5.4.8 Chapter 5.4.6
Resource utilisation and costs	Resource use and costs accounted for in the model are drug acquisition costs, administration costs, monitoring costs, adverse events costs, costs of subsequent treatment, and terminal care costs.	KEYNOTE-087 and Cheah[7] studies and published sources were used when they provided estimates of resource use and costs. This approach had been validated by expert opinion in previous submissions. Sources used are the eMIT,[37] BNF, the KEYNOTE-087 and Cheah[7] studies and studies used in TA462.	Chapter 5.5
Discount rates	Discount of 3.5% for utilities and costs	As per NICE reference case	Table 54
Sub groups	Not applicable		
Sensitivity analysis	Both DSA and PSA were performed as well as scenario analyses	As per NICE reference case	Chapter 5.8

Source: CS[1]

AE = adverse events; alloSCT = Allogeneic Stem Cell Transplant; autoSCT = Autologous Stem Cell Transplant; BNF = British National Formulary; BV = Brentuximab Vedotin; BSC = best supportive care; CS = company submission; DSA = deterministic sensitivity analysis; eMIT = electronic market information tool; OS = overall survival; PFS = progression-free survival; PSA = probabilistic sensitivity analysis; RRcHL = relapsed or refractory classical Hodgkin Lymphoma; SLR = systematic literature review; SoC = standard of care.

5.2.1 NICE reference case checklist (TABLE ONLY)

Table 5.2: Summary of the company's economic evaluation (with signposts to CS): NICE reference case checklist

Elements of the economic evaluation	Reference Case	Included in submission	Comment on whether *de novo* evaluation meets requirements of NICE reference case
Population	As per NICE scope[4]	Yes	
Comparator(s)	Therapies routinely used in the National Health Service (NHS), including technologies regarded as current best practice	Partly	BSC is only used in a scenario
Type of economic evaluation	Cost effectiveness analysis	Yes	
Perspective on costs	NHS and Personal Social Services (PSS)	Yes	
Perspective on outcomes	All health effects on individuals	Yes	
Time horizon	Sufficient to capture differences in costs and outcomes	Partly	Time horizon of 40 years, used in the base-case, does not capture all relevant costs and effects (illustrated in CS scenario analysis 5[1])
Synthesis of evidence in outcomes	Systematic review	Yes	SLR and naïve treatment comparison.
Measure of health effects	Quality adjusted life years (QALYs)	Yes	
Source of data for measurement HRQoL	Described using a standardised and validated instrument	Yes	
Source of preference data for valuation of changes in HRQoL	Time-trade off or standard gamble	Yes	
Discount rate	An annual rate of 3.5% on both costs and health effects	Yes	
Equity weighting	An additional QALY has the same weight regardless of the other characteristics of the individuals receiving the health benefit	Yes	
Sensitivity analysis	Probabilistic modelling	Yes	

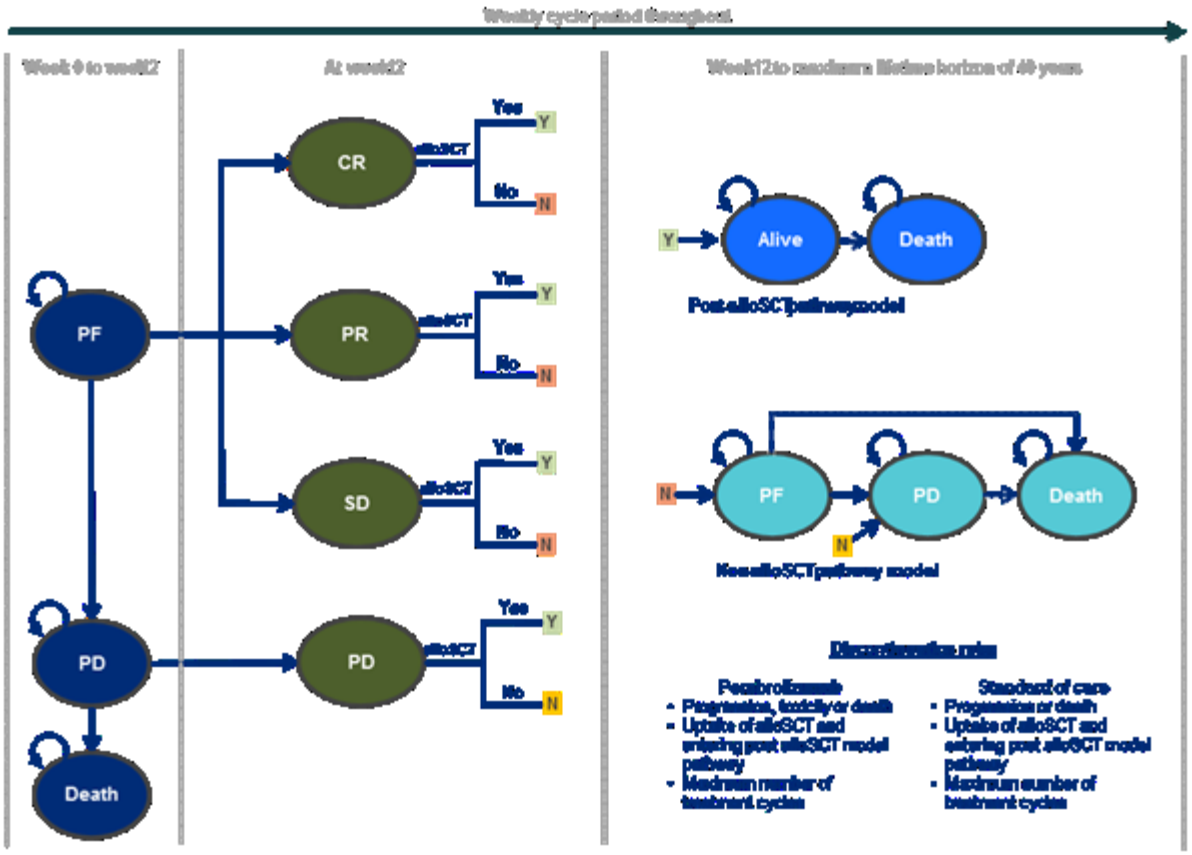
Source: CS [1]
NHS = National Health Service; PSS = Personal Social Services; QALY = quality-adjusted life year

5.2.2 Model structure

A de novo cohort state transition model was developed with health states based on response, uptake of alloSCT, and survival. This approach was adopted as it is expected that pembrolizumab monotherapy will result in higher response rates than SoC and hence will be used as a "bridge" to alloSCT. More specifically, pembrolizumab aims to control the disease, and if possible, elicit a disease response that enables alloSCT. The model has a time horizon of 40 years, weekly cycle length and applies a half-cycle correction.

The model structure consists of a short-term component (first 12 weeks), a subsequent decision tree element (at 12 weeks) to determine the proportion of patients transiting to alloSCT (conditional upon response at 12 weeks) and a long-term component (after the first 12 weeks) separately for patients who had alloSCT and patients who did not have alloSCT at 12 weeks (See Figure 5.1).

Figure 5.1: Model structure



Source: CS Figure 13[1]

Short term model structure (pre-12 weeks)

A partitioned survival approach is used for the first 12 weeks with three health states:

1. Progression free;
2. Progressed disease;
3. Death

Decision tree element (at 12 weeks)
After 12 weeks, the progression free proportion is subdivided into proportions of patients with complete response, partial response and stable disease. Here, patients with non-evaluable response status are categorised as having stable disease. Subsequently, depending on this response status, the proportion of patients continuing to alloSCT is calculated (i.e. patients with complete response have a higher probability of receiving alloSCT than patients with partial disease or stable disease). The company assumed that none of the patients with progressed disease will continue to alloSCT (see Section 5.2.6 for more details).

Long-term model structure (post-12 weeks) separately for alloSCT and non-alloSCT treatment
After the 12-week decision tree element, the cohort is split into patients who did and did not receive alloSCT.

Patients who did not receive alloSCT at 12 weeks will not be able to receive alloSCT for the remainder of the model time horizon. Further, the long-term costs and effects for this group are modelled using three health states consistent with the short-term model structure (for the first 12 weeks):

1. Progression free (patients who did not have alloSCT, progression, or died in the first 12 weeks);
2. Progressed disease;
3. Death

The long-term costs and effects for patients who did receive alloSCT at 12 weeks is modelled using two health states:

1. Alive (patients who did not have progression in the first 12 weeks and did receive alloSCT at week 12);
2. Death

Post-alloSCT survival was assumed to be independent of prior therapy (i.e. equal for patients who initially received pembrolizumab monotherapy and SoC). Moreover, the company justified not considering post-alloSCT progression in the model structure by claiming that the consequences of post-alloSCT progression are incorporated in the post-alloSCT utilities and costs.

ERG comment: The ERG notes the following issues regarding the model structure used by the company: (a) in the model it is only possible to have alloSCT 12 weeks after treatment start, (b) the assumption that alloSCT would be performed immediately after response; (c) neglecting a progression health state after alloSCT.

(a) The model structure only allows patients to have alloSCT at 12 weeks after starting pembrolizumab or SoC. No justification was provided for why this simplifying approach was adopted. This is of particular concern given that one of the main goals of pembrolizumab is to enable alloSCT and hence this should be reflected in the model as accurately as possible. Therefore, the ERG requested an analysis removing this assumption (i.e. incorporating a continuous alloSCT probability). However, in response to clarification question B4d, the company stated that they could not perform such an analysis given that 1) they believed that alloSCT data from KEYNOTE-087 are not reflective of UK clinical practice and; 2) they did not have Kaplan-Meier data for time-to-alloSCT from Cheah et al.[7]

Furthermore, the 12-week timepoint is questionable. It was selected based on a UK clinician survey and the company stated (clarification question B4a) that this timepoint is an accurate representation of the timing of the decision to transplant. The company recognised that response might be obtained later than week 12, but believed the assumption that these ‘later responders’ would not be considered for alloSCT

to be conservative. The ERG is not convinced that this statement is correct given that this was not appropriately explored by the company and it is unclear how many 'later responders' exist for both pembrolizumab and SoC.

The company's approach is furthermore inconsistent with the approach taken in TA462.[6] The company refers to TA462[6] on multiple occasions to highlight the similarities. This includes similarities regarding the mechanism of action of pembrolizumab and nivolumab stating that both may act as therapy to enable alloSCT. Therefore, it is questionable why the company opted to use a different model structure than in TA462.[6] In TA462,[6] alloSCT is assumed to be performed at six months.

(b) Another related concern is that the company assumes an immediate procedure at the 12-week time point. The company's model structure estimates the proportion of patients undergoing alloSCT based on response at week 12 after starting pembrolizumab or SoC and alloSCT would be performed immediately. This, however, neglects the time required to identify a donor and schedule the procedure. The lag is estimated to be on average [redacted] weeks from eligibility decision to the actual performing of alloSCT (given the company stated treatment is stopped on average [redacted] weeks prior to alloSCT). Hence, the decision to perform alloSCT might be made at 12 weeks, the actual procedure might be performed between 12 and 24 weeks (response to clarification question B4a). This would also be more consistent with TA462[6] wherein it is stated that "*Based on CheckMate 205 and the published literature, it has been assumed that a proportion of eligible patients with an adequate response will receive alloSCT at six months.*" This entails that alloSCT in the present model is performed earlier than would be expected in clinical practice. Hence, the post-alloSCT benefits (e.g. lower mortality probability and higher quality of life) are applied earlier. Given that the proportion of patients proceeding to alloSCT is higher for pembrolizumab than for SoC, this is most likely not a conservative assumption.

(c) As highlighted by the company (CS section 5.2.2), one of the main criticisms on partitioned survival models (recent Decision Support Unit report[38]), is that OS, a key driver of QALY gains in advanced oncology, is modelled independently of an underlying disease model. Hence, partitioned survival models might result in inappropriate extrapolations. This critique is applicable to the long-term post-alloSCT component of the model in which disease progression is not considered despite Lafferty et al[21] reporting a progression free survival one-year post-alloSCT of only 54%. Given that post-alloSCT survival is modelled independently of an underlying disease model, this likely biases the long-term extrapolations, in favour of pembrolizumab. This is also inconsistent with TA462 in which post-alloSCT progression was incorporated.

5.2.3 Population

According to its marketing authorisation, pembrolizumab monotherapy is indicated for the treatment of adult patients with relapsed or refractory classical Hodgkin lymphoma (RRcHL) who have failed autoSCT and brentuximab vedotin (BV) (cohort 1), or who are transplant ineligible and have failed BV (cohort 2). In line with this marketing authorisation and the final scope issued by NICE,[4] two distinct populations are considered in the cost effectiveness model:

- Cohort 1: RRcHL who have failed autoSCT and BV
- Cohort 2: RRcHL who are autoSCT ineligible and have failed BV

See Table 5.3 for the baseline characteristics for cohorts 1 and 2 (from the main evidence sources considered in the model).

Table 5.3: Baseline characteristics from the main evidence sources considered in the model

Characteristic		KEYNOTE-087, Cohort 1	KEYNOTE-087, Cohort 2	Cheah et al. (2016),[7] Cohorts 1 and 2
Treatment		Pembrolizumab 200mg		Mix of therapies including chemotherapy, and investigational agents
Number of patients		69	81	97 or 89
Age (median)		34.0	40.0	32
Age >45 (%)		25%[γ]	42%[γ]	14 (14%)
Female (%)		33 (47.8%)	38 (46.9%)	46 (47%)
ECOG	0	29 (42.0%)	44 (54.3%)	33 (41%)
	1	39 (56.5%)	37 (45.7%)	44 (54%)
	2	1 (1.4%)	0 (0.0%)	3 (4%)
Stage	1	NR	NR	2 (3%)
	2	NR	NR	25 (30%)
	3	NR	NR	18 (21%)
	4	NR	NR	39 (46%)
Baseline B symptoms		22 (31.9%)	26 (32.1%)	7 (8%)
Haemoglobin <105 g/l		35%	27%	18 of 51 (35%)
Lymphocytes <0.6 × 10^9/l		19%	15%	19 of 46 (41%)
White cell count >15 × 10^9/l		9%	17%	4 of 82 (5%)
Albumin <40 g/l		48%	49%	23 of 82 (28%)
Any extranodal site		56%	41%	31 of 88 (35%)
Maximum tumour diameter ≥4 cm		49%	42%	18 of 69 (26%)
Bulky Lymphadenopathy		5 (7.2%)	11 (13.6%)	15 (37%)
Bone marrow involvement		3 (4.3%)	5 (6.2%)	NR
Disease status - relapse		46 (66.7%)	24 (29.6%)	NR
Disease status – refractory		23 (33.3%)	57 (70.4%)	NR
Previous BV therapy		69 (100.0%)	81 (100.0%)	89 (100%)
Prior autoSCT		69 (100.0%)	0 (0.0%)	66 of 97 (68%)
Prior radiation		31 (44.9%)	21 (25.9%)	NR
Median no. of prior line of therapy		4	4	3

Source: CS, Table 53, CS Appendix 8, and Tables 1 and 2 in Cheah et al. 2016
BV = Brentuximab vedotin; ECOG = Eastern Cooperative Oncology Group; N/A = Not applicable; NR = Not reported; SCT = Stem cell transplant

ERG comment: The populations described by the company are consistent with the final scope issued by NICE for this appraisal,[4] but one concern relates to the use of a mixed population comparator.

For KEYNOTE-087, the company was able to distinguish between patients who did and did not receive autoSCT (i.e. cohort 1 and 2 respectively). For the study by Cheah et al.,[7] the company did not have

access to the individual patient level data and hence was unable to make this distinction. Hence, the mixed population from Cheah et al.,[7] including both patient groups that did and did not receive autoSCT, was used for both cohort 1 and 2. Given that the majority of patients (68%) in the study by Cheah et al.[7] did receive autoSCT, this mixed population is more reflective of cohort 1. Additionally, the Cheah et al.[7] population is more comparable with KEYNOTE-087 cohort 1 than with cohort 2 in terms of patient characteristics (see for instance baseline age, ECOG, haemoglobin and white cell count). For other baseline variables (e.g. baseline B symptoms, lymphocytes, albumin, extranodal sites, tumour diameter >=4cm, and bulky lymphadenopathy) the Cheah et al.[7] population differs from both KEYNOTE-087 cohorts.

In response to clarification question B9 the company states that, based on clinical opinion, cohort 2 represents a higher risk group that is likely to progress more quickly compared with cohort 1. If this is the case, using the mixed population from Cheah et al.[7] in the naive comparison likely resulted in comparisons of pembrolizumab with SoC that may be favourable and non-favourable for pembrolizumab in cohorts 1 and 2 respectively.

5.2.4 Interventions and comparators

Pembrolizumab monotherapy is implemented as per its EMA Summary of Product Characteristics (SmPC) posology and method of administration for RRcHL (i.e. administered intravenously at a fixed dose of 200mg over 30 minutes every 3 weeks [Q3W]). The company assumed that in the model pembrolizumab monotherapy will be provided for a maximum of 24 months (35 cycles).

The NICE scope specifies the following comparators:

- Single or combination chemotherapy including drugs such as gemcitabine, vinblastine and cisplatin;
- Best supportive care (BSC).

The company only considered "standard of care" (SoC) as comparator in its base-case. SoC as considered by the company consists of the following regimens:

- chemotherapy (see CS Table 88 for the included treatments);
- bendamustine or;
- investigational agents.

The distribution of patients among these regimens was based on the distribution observed in Cheah et al (2016)[7] (see CS Table 88).

The company also presented a scenario analysis, in which BSC was added as a comparator. The company justified this deviation from the scope (i.e. not including BSC in its base-case) by stating they believed BSC use to be minimal as eligible patients are likely to receive therapy whenever feasible.

ERG comment: The ERG has concerns regarding (a) the exclusion of BSC from the base-case, (b) the recent recommendation of nivolumab in part of this population, which is not reflected in the analysis, (c) the assumption that pembrolizumab treatment stops at 24 months, and (d) the inclusion of investigational agents in the comparator.

(a) Regarding the inclusion of comparators, the ERG wishes to highlight that BSC is not incorporated in the CS base-case (inconsistent with the scope), but only presented in a scenario analysis.

(b) Moreover, nivolumab was recommended by NICE in part of this population (cohort 1). Nevertheless, NICE (personal communication with [redacted]) suggested that it would be

inappropriate to include nivolumab as a new comparator given it is still within the 90-day implementation period and hence is not considered established practice.

(c) The assumption that pembrolizumab monotherapy will be stopped after 24 months (35 cycles) is inconsistent with the SmPC but in line with the KEYNOTE-087 protocol. It is unclear whether pembrolizumab, in UK clinical practice, would also be provided for a maximum of 24 months. The company explored the impact of this assumption in a scenario in which patients continue treatment after 24 months. This scenario increased the CS base-case ICER for both cohorts (response to clarification question B13).

(d) Finally, the company uses the total population from Cheah et al (2016),[7] including patients that received investigational agents. Given that excluding patients that received investigational agents might result in a selected patient sample, the ERG believes this approach is reasonable. Moreover, the appropriateness of using the patients that used investigational agents in the Cheah et al (2016)[7] study was discussed in the final appraisal determination (FAD) of TA462. The committee preferred to use the overall population from Cheah et al (2016)[7] given that it considered that *"selectively excluding potentially the fittest patients from the Cheah dataset could bias the results of the indirect treatment comparison more than including some treatments that may not be used in UK current practice"*.[6]

5.2.5 Perspective, time horizon and discounting

The model adopts the perspective of the NHS and Personal and Social Services (PSS) in England and Wales. The cycle length is one week to account for the length of treatment cycles. A time horizon of 40 year was adopted to capture all relevant costs and outcomes. All costs and utilities were discounted at a rate of 3.5% per year.

ERG comment: The ERG considers the adopted perspective and discounting to be appropriate for this appraisal. The time horizon of 40 year might be considered suboptimal given that CS scenario 5 (CS section 5.8.3) suggests that this time horizon is insufficient to capture all costs and outcomes. Therefore, the ERG preferred to use a 50-year time horizon in its base-case.

5.2.6 Treatment effectiveness and extrapolation

Treatment effectiveness for pembrolizumab was primarily based on the KEYNOTE-087 study.[10] The only comparator in the company's base-case was SoC. The primary data source for the SoC comparator was the Cheah et al (2016) study.[7] The company performed a naïve indirect treatment comparison to derive hazard ratios for OS and PFS and response rates at week 12. A MAIC was also performed and results are shown in the company's scenario analysis. Both KEYNOTE-087 cohorts were compared with the Cheah et al (2016) study.[7] In a scenario analysis, the company explored BSC as a comparator. Because no data were available to inform this comparator, the efficacy of SoC was used (CS p. 149).

Due to the company's model structure, treatment effectiveness and time to treatment discontinuation (TTD) were estimated for the pre-12-week period and for the post-12-week period separately. Parametric models were fitted to data from KEYNOTE-087 to estimate OS and PFS for patients receiving pembrolizumab in the pre-12-week period. To inform the decision tree element at week 12, response rates from KEYNOTE-087 were used, as well as two clinician surveys to inform estimates of probability of alloSCT conditional on response status (i.e. complete response, partial response, stable disease). For the post-12-week period, treatment effectiveness depended on whether patients received alloSCT or not. Mortality post-alloSCT was based on Lafferty et al[21] and mortality for patients who did not receive alloSCT was based on Cheah et al.[7] PFS for patients who did not receive alloSCT was

estimated from KEYNOTE-087. The company justified this inconsistency by stating that survival data from KEYNOTE-087 were immature.

TTD for the pre-12-week period was assumed to be equivalent to PFS. TTD for the post-12-week period was estimated directly from KEYNOTE-087. Furthermore, TTD for SoC was assumed equivalent to PFS in Cheah et al for pre- and post-12 weeks.[7]

Table 5.4 presents an overview of use and justification of all parametric models for PFS and OS extrapolations in the two periods, with more detail provided in the following sections.

Table 5.4: Overview of parametric models used for extrapolating OS and PFS in company model

	Parametric model used in company base-case	Best statistical fit? (if No: which one?)	Other justification provided?	Alternative explored in company scenario analysis?	Source used for pembrolizumab
Cohort 1					
Pre-12 weeks PFS	Log-logistic	Yes	None	No	KEYNOTE-087
Pre-12 weeks OS	Lognormal	No (exponential)	Predicted highest mortality	No	KEYNOTE-087
Post-12 weeks PFS	Exponential	Yes	None	No	KEYNOTE-087
Post-12 weeks (non-alloSCT) OS	Constant transition probability estimated from median OS		No KM estimates available from Cheah, KEYNOTE-087 data too immature	Yes, KEYNOTE-087 data were explored in scenario analysis	Cheah et al (2016)[7]
Post-12 weeks (-alloSCT) OS	Weibull	No (gen gamma)	Gen gamma predicted an infinite hazard beyond 150 months and had to be adjusted, thereby under-estimating the survival benefit; AIC/BIC scores were relatively similar; ERG in TA462 considered lognormal and Weibull most clinically plausible	Lognormal	Lafferty et al (2017)[21]
Pre-12 weeks TTD	Same as pre-12 weeks PFS				KEYNOTE-087
Post-12 weeks TTD	Exponential	Yes	Maintained consistency with post-12 week PFS	No	KEYNOTE-087

	Parametric model used in company base-case	Best statistical fit? (if No: which one?)	Other justification provided?	Alternative explored in company scenario analysis?	Source used for pembrolizumab
Cohort 2					
Pre-12 weeks PFS	Generalised gamma	Yes	None	Weibull	KEYNOTE-087
Pre-12 weeks OS	Exponential	No (lognormal)	None	No	KEYNOTE-087
Post-12 weeks PFS	Exponential	No (gen gamma)	The last drop in PFS was not considered informative, due to small patient numbers at risk	Gompertz	KEYNOTE-087
Post-12 weeks (non-alloSCT) OS	Constant transition probability estimated from median OS		No KM estimates available from Cheah, KEYNOTE-087 data too immature	Yes, KEYNOTE-087 data were explored in scenario analysis	Cheah et al (2016)[7]
Post-12 weeks (-alloSCT) OS	Same as for cohort 1				Lafferty et al (2017)[21]
Pre-12 weeks TTD	Same as pre-12 weeks PFS				KEYNOTE-087
Post-12 weeks TTD	Same as for cohort 1				KEYNOTE-087
OS = overall survival; PFS = progression free survival; TTD = time to treatment discontinuation					

ERG comment: The ERG's general comments on treatment and relative treatment effectiveness used in the model relate to (a) inconsistency in the choice of data sources prompted by the immaturity of OS data in KEYNOTE-087, (b) the lack of BSC as a comparator, (c) the use of a naïve indirect comparison and (d) the use of differential parametric models for the pre- and post-12-week periods.

(a) For the post-12 weeks period, the company deviated from their main data source and used the Cheah and Lafferty et al studies to inform mortality for patients without and with alloSCT respectively. This was justified by the company by stating that KEYNOTE-087 OS data were too immature to be used. [redacted] and the ERG considers that these may be informative for the present analysis.

(b) The lack of BSC as a comparator is non-compliant with the scope. The company justified this stating that there were no data to inform this comparison, and provided a conservative scenario analysis in which the effectiveness of BSC was assumed equivalent to that of SoC.

(c) The company's argument for preferring the naïve treatment comparison to minimise data loss (see CS p 149) is plausible in the context of small sample sizes. The MAIC is deemed to introduce systematic error, due the limited availability of prognostic variables. The ERG therefore maintains the naïve

comparison in its base-case and the MAIC is explored in scenario analysis. The naïve comparison favours SoC.

(d) The artificial 12-week time point necessitated the fitting of differential curves to the pre- and post-12-week periods. This leads to loss of data introducing further uncertainty in the extrapolation.

5.2.6.1 Pre-12 weeks: PFS and OS

PFS pre-12 weeks

PFS pre-12 weeks was modelled based on the entire observed data set from KEYNOTE-087 beginning in week 0 to the end of study follow-up. The company justified this by stating that there was only a small number of events occurring in the first 12 weeks.[10] The log-logistic model was deemed to best represent PFS for cohort 1 and the generalised gamma for cohort 2 (based on best statistical fit). The company stated that in cohort 2, the generalised gamma over-predicted the number of patients in the progression-free survival health state and the company explored the Weibull in a scenario analysis, claiming that it would result in fewer patients in the progression-free health state at 12 weeks.

Relative effectiveness was based on the naïve treatment comparison.

OS pre-12 weeks

OS pre-12 weeks was also based on KEYNOTE-087. With very few events, there was no meaningful difference between the different parametric models in terms of statistical fit and the company selected the lognormal model for cohort 1, which predicted the highest mortality but did not have the best statistical fit. For cohort 2, [redacted]. The company chose the exponential model, without providing appropriate justification.

The company assumed that patients treated with SoC would follow the same OS curve as patients receiving pembrolizumab.

ERG comment: The ERG wishes to highlight a few caveats with the company's pre-12 weeks analysis, including (a) the fitting of parametric models for the pre-12-week period using the entire study data, and (b) the poor fit of models for OS in both cohorts, which produces artificially lowered LYs and counter-intuitive results.

(a) Only very few events occurred in the first 12 weeks of the KEYNOTE-087 study. For example, for PFS, more than [redacted] of patients in cohort 1 and approximately [redacted] of patients in cohort 2 were still progression-free at 12 weeks. The fitted models were estimated using the entire study data from week 0 to end of study follow-up, which may have led to the fitted curves being more influenced by the post-12-week period than the pre-12-week period. This is exacerbated for PFS in cohort 2. This is because the KM estimates show that there is a significantly increased rate of progression starting at 11 weeks. This sudden drop, as well as having the parametric models fitted to the entire study data, results in most of the curves not providing a good fit. Furthermore, the scenario analysis using the Weibull over-predicts patients in the progression-free health state even more than the base-case generalised gamma, contrary to the claims of the company. This analysis is therefore disregarded by the ERG, as the only rationale for scenario analysis using the Weibull for PFS in cohort 2 was that it over-predicted PFS to a lesser extent than the generalised gamma. The ERG therefore considers the company's adopted approach of deriving pre-12 weeks PFS and OS estimates from the entire study data as questionable.

(b) For both cohorts, the company chose the pre-12-week OS models that predicted the highest mortality at 12 weeks, disregarding statistical fit (lognormal for cohort 1 and exponential for cohort 2). This likely

produces an artificially lowered number of life-years (LYs) gained, however, it may be worth noting that the company's economic model overall predicts LYs that were considered by the company to be high[1] due to the inclusion of investigational agents in Cheah et al.[7] The combination of using the generalised gamma for PFS and the exponential for OS in cohort 2 also resulted in the crossing of PFS and OS curves in the model (first PFS > OS, then PFS < OS). The company remedied this by choosing whichever was smaller in the simulation of PFS. The ERG preferred to use the model with the best statistical fit in their base-case. This, however, did not solve the problem of crossing OS and PFS.

5.2.6.2 At 12 weeks: response rates and alloSCT probabilities

Response rates at 12 weeks

The distribution across the response states of complete response (CR), partial response (PR), stable disease (SD) and progressed disease (PD) was based on the observations from the KEYNOTE-087 study. The company only presented the patient numbers for cohort 1 in Table 62 of the CS,[1] but corrected this in their response to the clarification letter (see Table 5.5).[10] The company furthermore highlighted in response to the clarification letter that all patients with a non-evaluable response status were assumed to have SD, and presented response rates in comparison with model predictions (Table 5.6).

Response rates at 12 weeks for SoC were based on odds ratios for response derived from the naïve treatment comparison.

Table 5.5: Response rates derived from KEYNOTE-087

Response	n	N
Cohort 1		
CR	[redacted]	[redacted]
PR	[redacted]	[redacted]
Cohort 2		
CR	[redacted]	[redacted]
PR	[redacted]	[redacted]
Source: Response to clarification letter[10]		

Table 5.6: Response rates and model predictions

Status	Table 19 of submission (cohort 1) N (%)	Model predictions (cohort 1) %	Table 19 of submission (cohort 2)	Model predictions (cohort 2)
Complete response	[redacted]	15.94%	[redacted]	8.6%
Partial response	[redacted]	42.0%	[redacted]	43.2%
Stable disease	[redacted]	36.9% *(~27.5%+8.7%)*	[redacted]	38.9% (~18.5% + 8.6%)
Non-evaluable	[redacted]	Not reported (combined in stable disease)	[redacted]	Not reported (combined in stable disease)
Progressed disease	[redacted]	4.10%	[redacted]	7.9%

Status	Table 19 of submission (cohort 1) N (%)	Model predictions (cohort 1) %	Table 19 of submission (cohort 2)	Model predictions (cohort 2)
Death	Not reported	1.04%	Not reported	1.22%
Source: Response to clarification letter[10]				

AlloSCT rates conditional on response

The probabilities of having an alloSCT conditional on response status were elicited through two clinician surveys, one performed by the company (referred to here as the MSD survey) and one performed within the course of previous TA462 (referred to here as the BMS survey).[6] The company stated that it was necessary to use the intermediate step of applying a probability of alloSCT based on response status, because it was not appropriate to use the KEYNOTE-087 study data on time to alloSCT directly. This was justified by a smaller proportion of patients ([redacted]) in the KEYNOTE-087 study receiving alloSCT,[1] compared with UK practice, although no data for UK practice, apart from the survey data, were presented. The KEYNOTE-087 study data on alloSCT were therefore not used directly to inform the present model. The company also stated that [redacted] of UK patients in cohort 1 and [redacted] in cohort 2 received alloSCT (CS p160)[1] but corrected this in their response to the clarification letter to be [redacted] for cohort 1 and [redacted] for cohort 2.[10]

The MSD clinician survey drew on opinions from 16 clinicians from the UK who were asked the proportion of patients they would expect to proceed to alloSCT conditional on response to treatment, which could be CR, PR, SD or PD. The results of this survey were combined with the results from the BMS survey by taking a simple, unweighted average of the means (Table 5.7). The company stated that it disregarded clinicians' responses indicating that some patients in the progressed disease health state could be eligible for alloSCT (a mean of [redacted] according to the company's slides in REF pack 1)[39] following further discussions with UK clinicians on this topic and stating that this was not thought to be standard UK clinical practice. However, in the KEYNOTE-087 study, [redacted] patients were in the progressed disease health state when they received alloSCT, albeit none of them from the UK.

For cohort 2, the same rates were assumed as for cohort 1, but some clinicians suggested that alloSCT rates in that population might be even higher than in cohort 1 due to the unmet need in this population.

The same alloSCT probabilities conditional on response status were adopted for both pembrolizumab and SoC.

Table 5.7: AlloSCT rates conditional on response

	MSD Mean[40]	**Alternative Mean**	**Overall Mean**	**SE**
CR	56.79%	[redacted]	[redacted]	[redacted]
PR	43.93%	[redacted]	[redacted]	[redacted]
SD	18.36%	[redacted]	[redacted]	[redacted]
Source: CS Table 64				

ERG comment: The ERG's comments include that (a) patients with a non-evaluable response status being considered to have SD inflates the proportion of patients in this health state, (b) the omission of

the survey result that patients with progressed disease could still be eligible for alloSCT is non-conservative, and (c) the combination of the MSD and BMS survey results may introduce bias.

(a) The proportions in the SD state in both cohorts in the model are significantly larger than those observed in the KEYNOTE-087 study, as reported in Table 19 of the CS. This is a result of patients with non-evaluable response status being moved into the SD state. In response to the clarification letter,[10] the company provided an overview of model predictions of response status compared with the KEYNOTE-087 data (shown in Table 5.8). It can be seen that the model may over-predict the proportions in the SD state, but this is likely a conservative assumption.

Table 5.8: Comparison of response status in model and KEYNOTE-087

Status	Table 19 of submission (cohort 1) N (%)	Model predictions (cohort 1) %	Table 19 of submission (cohort 2)	Model predictions (cohort 2)
Complete response	[redacted]	15.94%	[redacted]	8.6%
Partial response	[redacted]	42.0%	[redacted]	43.2%
Stable disease	[redacted]	36.9% *(~27.5%+8.7%)*	[redacted]	38.9% (~18.5% + 8.6%)
Non-evaluable	[redacted]	Not reported (combined in stable disease)	[redacted]	Not reported (combined in stable disease)
Progressed disease	[redacted]	4.10%	[redacted]	7.9%
Death	Not reported	1.04%	Not reported	1.22%
Source: Response to clarification letter				

(b) Patients with PD were assumed to not get alloSCT, despite the MSD survey results indicating otherwise ([redacted] of patients with PD would get alloSCT). In response to the clarification letter,[10] the company explained that based on feedback from UK clinicians, it is not UK standard practice that patients in the PD state would receive alloSCT. The company furthermore provided data from KEYNOTE-087, where none of the [redacted] UK patients who underwent alloSCT (in cohorts 1 and 2) had PD prior to alloSCT, but [redacted] was non-evaluable. The ERG was concerned about this argumentation. First, the MSD survey was performed with UK expert clinicians only and it was not explained why the company considered it appropriate that discussions with a number of UK clinicians overrode the survey results. Furthermore, the [redacted] UK patients from KEYNOTE-087 who underwent alloSCT are likely too few to be representative. The company did provide a scenario analysis enabling alloSCT in PD patients, which resulted in increases in the ICER. The ERG adopted the MSD survey results with probabilities for alloSCT in PD patients in its base-case.

(c) The ERG considers the combination of the MSD and BMS surveys as problematic: for one, the company stated that the TA462 committee had deemed the Cheah et al[7] estimates of 66% of responders receiving alloSCT as too high for the UK. However, it can be seen from Table 5.7 that estimated proportions of patients receiving alloSCT from the MSD survey were lower than those from the BMS survey. Hence, when both surveys are combined, according to Table 5.7, the alloSCT rates used in the CS for the CR status are even higher than the Cheah et al[7] estimates, and even when the mean for PR and CR is taken, the resulting alloSCT rates for responders ([redacted]) are not significantly lower than those in Cheah et al.[7] Given that the company's estimation of alloSCT rates based on their own MSD survey would have resulted in lower alloSCT rates for partial and complete responders ([redacted]) compared to

Cheah et al[7], the ERG considers the use of the MSD survey data alone to be more in line with the TA462 committee preferences. The committee conclusion on the BMS survey also entailed the following comment: "the committee also heard that recent NHS referrals for allogeneic stem cell transplant were lower than those reported in the [BMS] survey." It is therefore not clear to the ERG why the company opted to combine the MSD and BMS surveys. This is of particular concern given that the company accepts that *"it is possible for both surveys to have included the same clinical experts"*.[10] It is the ERG's view that bias induced by double-counting of certain experts' opinions cannot be ruled out. The company, in response to the clarification letter, provided a scenario analysis using alloSCT rates from the MSD survey only, which indicated that the ICERs increased. For reasons mentioned above, the ERG preferred to use the MSD survey only, instead of combining them with the BMS survey, in its base-case.

5.2.6.3 Post-12 weeks: patients not receiving alloSCT – PFS and mortality

PFS post-12 weeks

PFS post-12 weeks was estimated using only the observed data from KEYNOTE-087 beginning in week 12 to end of study follow-up.[10] The company stated that alloSCT events were not censored from the survival analysis of KEYNOTE-087 because it was not possible to censor them from the Cheah study either. The exponential distribution was used to estimate PFS post-12 weeks in cohorts 1 and 2. For cohort 1, this represented the model with the best statistical fit. For cohort 2, the exponential did not make the best statistical fit (the generalised gamma did) and it over-estimated PFS at the end of follow-up, however the company argued that the last drop in PFS was not considered particularly informative given the small patient numbers at risk (n=3) (see Figure 5.2). The Gompertz was considered in a scenario analysis.

The company assumed that the post-12 week HR and pre-12 week HR for PFS were equal for cohorts 1 and 2 estimated at [redacted] and [redacted], respectively, for the entire study period. The company justified this by stating that *"a large number of progression events occurred during the first 12 weeks of the SoC study."* and that *"Therefore, it was not possible to estimate a HR between the two treatments after 12 weeks."* (CS p 151) [1] The company concluded that *"a PFS HR from week 12 to end of follow-up could not be estimated given the low number of events post week 12 observed in Cheah"* (CS p 141). [1] In response to the clarification letter,[10] the company furthermore clarified that the HR for pembrolizumab versus SoC was derived using the entire follow-up period.

Figure 5.2: PFS (BIRC) cohort 2 from week 12 extrapolations

CS Figure 21[1]

Mortality pre-progression post-12 weeks

Mortality in the pre-progression health state post-12 weeks when patients did not receive alloSCT was assumed to be equal to general mortality estimates derived from UK life tables for both pembrolizumab and SoC.[41]

Mortality post-progression post-12 weeks

Because the number of patients was considered too small to support robust analysis of post-progression survival, the company used Cheah et al (2016)[7] to estimate post-progression mortality for both pembrolizumab and SoC. The weekly transition probabilities were obtained by converting median OS in Cheah assuming a constant hazard rate based on the exponential distribution. The obtained transition probability of 0.63% (per week) was replaced by background mortality when general mortality estimates obtained from UK life tables exceeded this probability.

The company implicitly assumed a HR = 1 for estimating mortality in the pre- and post-progression health states by using general mortality estimates for pembrolizumab and SoC for the pre-progression health state, and Cheah et al.[7] to inform transition probabilities from the post-progression state to the dead state.

ERG comment: The ERG's concerns relate to (a) uncertainty around extrapolating PFS post-12 weeks, (b) the assumption that patients in the pre-progression health state can only die from all-cause mortality, (c) the assumption that pre- and post-12 week HRs for PFS were equal, and (d) the immature OS data from KEYNOTE-087.

(a) For post-12 week PFS in cohort 1, the choice of the exponential distribution was based on best statistical fit. The Gompertz distribution had a statistical fit within two AIC points and the ERG therefore considered it informative to explore the use of this model in scenario analysis.

In cohort 2, the choice of the exponential distribution for post-12 week PFS is unclear. The generalised gamma distribution has the best statistical fit, followed by the Gompertz and exponential distributions (based on AIC and BIC respectively). Despite this, the company chose the exponential distribution, with the rationale that the small patient numbers at risk at the end of follow-up make the last drop less informative. The ERG considers clinical plausibility important but remains unconvinced that there was sufficient justification for ruling out the generalised gamma. Clinical expert opinion should have been used to validate this assumption. The ERG considers that the model with the best fit (generalised gamma) and second best fit (Gompertz) should be explored in scenario analysis. Results show that the choice of post-12 week PFS model in cohort 2 is very influential and that the company's choice of exponential favoured pembrolizumab.

(b) In the pre-progression health state, patients are assumed to die only from all-cause mortality. There was no indication provided for why this was clinically plausible and the ERG is uncertain about the impact of this assumption on model outcomes.

(c) The ERG considers the assumption that post-12 week HR and pre-12 week HR for PFS were equal to be questionable. The use of a constant HR lacks face validity because different parametric models pre- (log-logistic and generalised gamma in cohorts 1 and 2) and post-12 weeks (exponential in cohorts 1 and 2) were used. The company, upon request, provided a scenario analysis using a HR=1 for the post-12 week period, which increased the ICERs significantly. This should be viewed as a worst-case scenario. Given that the HRs were estimated based on the entire study data, the ERG maintains the HRs used by the company in its base-case.

(d) OS data for the entire study population of KEYNOTE-087 was deemed by the company to be too immature to provide robust extrapolations of survival.[7] Upon request, the company provided scenario analysis with post-12 weeks post-progression survival estimated based on KEYNOTE-087 instead of Cheah et al.[7], which decreased the ICERs. Because of the small number of post-progression events in KEYNOTE-087 ([redacted] in cohort 1, [redacted] in cohort 2),[10] the ERG agrees that these data are too immature to be used in the present analysis.

5.2.6.4 Post-12 weeks: patients receiving alloSCT - OS

OS estimates were obtained from a UK study consisting of 13 patients with classical Hodgkin Lymphoma who received alloSCT after three previous therapies (Lafferty et al, 2017).[21] The company stated that this was in line with previous TA462 on nivolumab for treating relapsed or refractory classical Hodgkin Lymphoma.[6] The company attempted to digitise the KM provided in Appendix 17 of the CS,[27] but resorted to manually adjusting the data because the unknown rate of censoring in the tail of the curve and the limited number of events prevented the company from reproducing patient level data. However, the company used the point estimates and AIC/BIC from TA462, and only used their own digitised version of the Lafferty KM data for the PSA.

The Weibull distribution was used to extrapolate OS beyond the available Lafferty et al data. The Weibull did not have the best statistical fit and, in fact, only came fifth according to the AIC/BIC criteria. However, the company justified their choice by stating that (1) the generalised gamma predicted an infinite hazard beyond 150 months and therefore had to be adjusted, thereby likely under-estimating the survival benefit expected in this population, (2) AIC/BIC scores were relatively similar (for example, AIC score of Weibull <3 points away from the AIC of the generalised gamma, which ranked

first in terms of AIC/BIC, (3) the ERG in TA462 considered the use of log-normal and Weibull models as more clinically plausible as they did not predict infinite survival, and (4) the company considered the Weibull more conservative than the lognormal. The lognormal was explored in the company's scenario analysis. Model predictions of the different models are shown in Table 5.9.

Table 5.9: Summary of the survival models (OS after alloSCT adjusted for all-cause mortality)

Item	Exponential	Weibull	Gompertz	Log-logistic	Log-normal	Generalised gamma	Lafferty 2017
Median (months)	53.13	64.62	266.78	58.41	61.86	87.39	--
Mean (months)	76.77	163.07	237.71	172.88	177.21	213.93	--
% at 1 year	85.73%	71.68%	63.33%	69.74%	70.01%	65.28%	64.17%
% at 2 years	73.39%	63.78%	55.90%	61.55%	61.93%	59.48%	53.47%
% at 5 years	53.77%	54.50%	53.58%	52.68%	53.33%	54.21%	53.47%
% at 10 years	21.09%	40.56%	52.90%	40.79%	41.77%	47.95%	--
% at 15 years	9.67%	34.13%	52.08%	35.78%	36.83%	45.43%	--
% at 20 years	4.43%	29.61%	50.80%	32.40%	33.45%	43.82%	--
% at 30 years	0.93%	23.46%	45.95%	27.88%	28.84%	39.63%	--
% at 40 years	0.20%	17.64%	34.77%	21.10%	21.83%	29.99%	--
Source: CS Table 69[1]							

ERG comment: The ERG has concerns about (a) the appropriateness of using Lafferty et al.[21] for estimating post-alloSCT OS and (b) that the company over-estimates OS in post-alloSCT patients.

(a) The ERG questioned the appropriateness of using Lafferty et al[21] for post-alloSCT survival, given that in KEYNOTE-087, [redacted] patients had an alloSCT compared with the 13 patients in Lafferty et al[21]. In response to the clarification letter,[10] the company explained that the KEYNOTE-087 study did not include the subsequent investigation of patients treated with pembrolizumab who were treated with a stem cell transplant. Furthermore, the company argued that OS data for the entire study population of KEYNOTE-087 were deemed to be too immature to provide robust extrapolations of survival and highlighted that Lafferty et al[21] was also used to inform TA462. Because Lafferty et al[21] is a very small study with questionable generalisability to the UK setting (see Section 4.2.3), its use means that there is substantial uncertainty around post-alloSCT survival, and alternative evidence was not explored.

(b) According to the company's Figure 3 in Appendix 17 of the CS,[27] (Figure 5.3) post-alloSCT survival is likely over-estimated. From this figure it appears that the company assumed no censoring after the last event until the end of the 5-year period. This results in an over-estimation of OS, as can be seen from the fitted curves that follow the plateau between 21 months and 5 years closely. It is unlikely that this plateau is a reflection of OS in patients post-alloSCT and the ERG considers it more likely that censoring occurred before the end of this 5-year period. The ERG acknowledges that there is uncertainty about the better approach, but notes that the company chose the approach that favoured pembrolizumab the most. The ERG therefore used the KM estimates from Figure 5.3 to reconstruct individual patient level data, allowing for censoring after the last event and before the end of the follow-up period, and used this in ERG scenario analysis, showing that the company's analysis significantly favoured pembrolizumab. The ERG's and the company's fitted curves are shown in Figure 5.4. As can be seen,

the ERG's approach gives less weight to the plateau in the tail of the Kaplan Meier curve than the company's approach.

Figure 5.3: KM estimates from Lafferty et al, as presented in company appendix 17

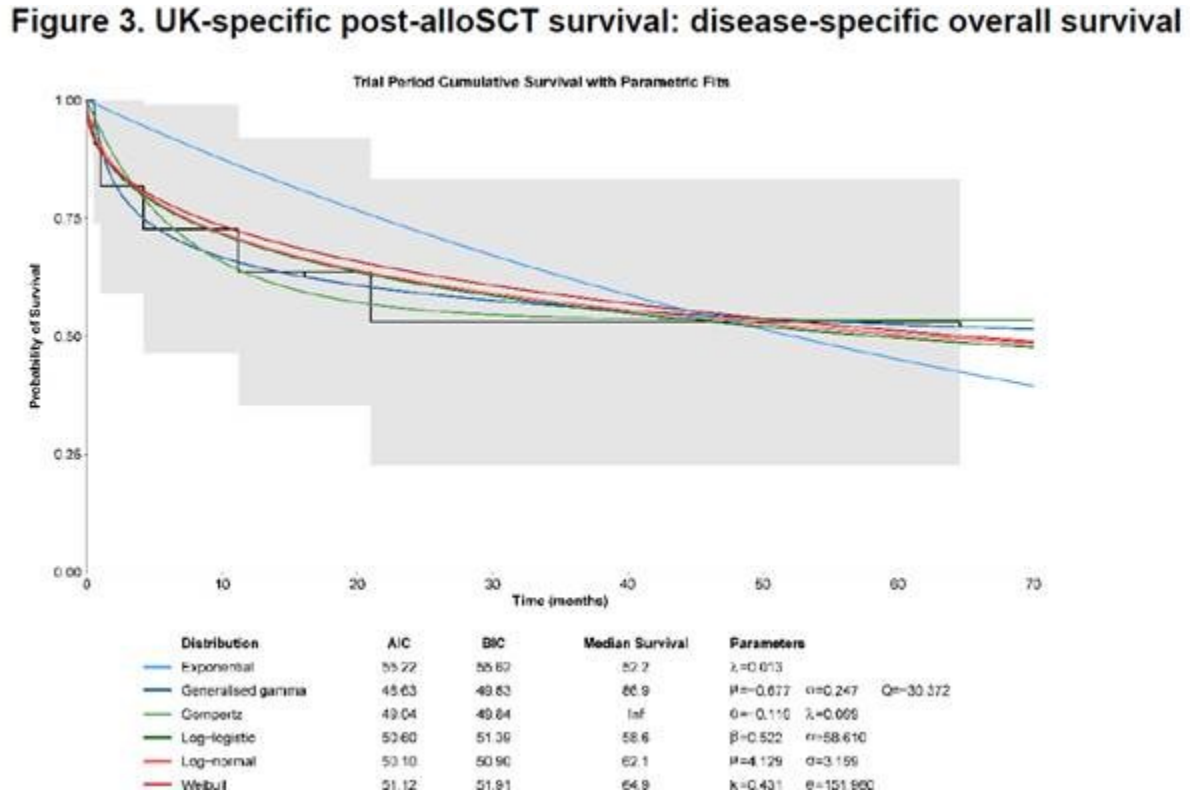
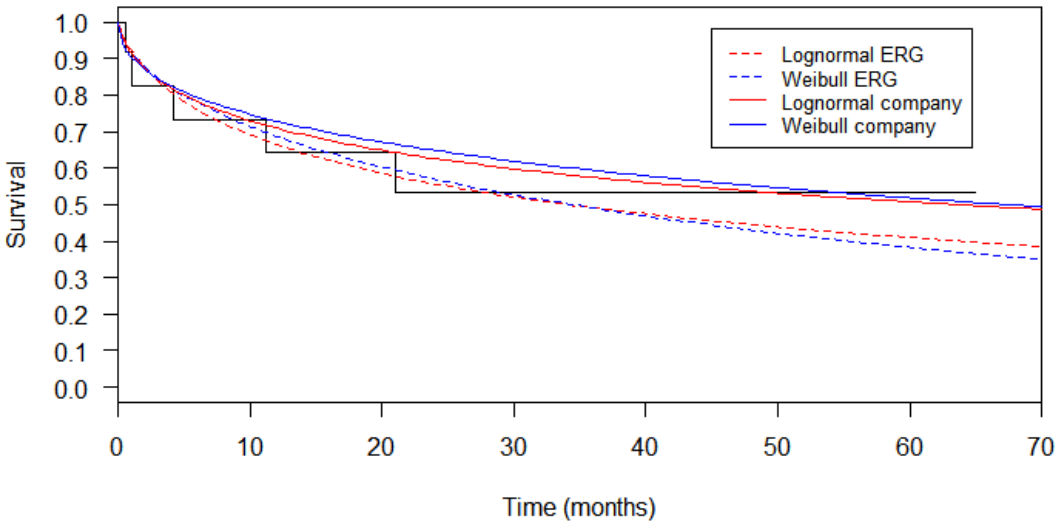


Figure 5.4 ERG's versus company's approach to estimating post-alloSCT OS based on Lafferty et al



5.2.6.6 Time to treatment discontinuation

Time to treatment discontinuation (TTD) pre-12 weeks
The company used PFS as a proxy for TTD for the pre-12 week period. No justification was provided.

TTD post-12 weeks
Treatment is discontinued for patients receiving alloSCT. For patients not receiving alloSCT, the company estimated time to treatment discontinuation for the post-12 week period using the TTD data available from KEYNOTE-087. PFS was not deemed an appropriate proxy because, on average, patients discontinued treatment before they progressed. The company postulated that this may be due to safety and tolerability and the impact of the design of KEYNOTE-087, which allowed study investigators to discontinue therapy if complete response had been achieved after at least six months of treatment. TTD is furthermore capped at 24 months in the company's model. The company justified this stating that this was in line with the stopping rule employed within the KEYNOTE-087 study.

For both cohorts 1 and 2, the exponential distribution was chosen, as it was the model exhibiting the best statistical fit and maintained consistency with the base-case PFS distribution.

For SoC, PFS was used as a proxy for TTD and this was justified by the lack of treatment discontinuation data from Cheah et al (2016).[7]

ERG comment: The ERG's concerns relate to (a) the inconsistency of using PFS as a proxy to TTD for the pre-12 weeks period and the comparator but not for TTD post-12 weeks, and (b) the capping of time to treatment discontinuation at 24 months in the model.

(a) For the pre-12 week period, PFS was used as a proxy to TTD. The company did not provide justification for this assumption. This means that the estimation of TTD is inconsistent between the pre- and post-12 weeks periods, and indeed with the comparator, for which PFS was used as a proxy.

(b) The company's assumption that treatment duration is capped at 24 months is not in line with the marketing authorisation. Upon request, the company provided a scenario analysis of continued treatment with pembrolizumab after 24 months, which showed that ICERs increased substantially for both cohorts.[10] This is possibly a pessimistic scenario, because effectiveness was based on KEYNOTE-087, in which the maximum treatment duration was 24 months. However, the ERG wishes to point out that the company's base-case might under-estimate the cost incurred with the use of pembrolizumab when a 24-months stopping rule is not enforced in clinical practice.

5.2.7 Adverse events

The company decided, in order to reflect best clinical practice, to incorporate the AEs that were included in the previous Hodgkin Lymphoma appraisal (TA462)[6]. Table 5.10 presents the grade 3+ AEs with an incidence of ≥0% in any study arm, that were incorporated as a one-off cost and disutility into the first cycle of the cost effectiveness model. The company assumed patients remaining on treatment beyond the first year to tolerate treatment well and therefore not to experience severe AEs. The company further assumed that investigational agents do not cause any AEs.

Table 5.10: Adverse event rates incorporated in the cost effectiveness model

Adverse Event	Pembrolizumab (cohort 1)	Pembrolizumab (cohort 2)	Chemotherapy	Bendamustine	SoC*
Anaemia	[redacted]	[redacted]	16.59%	13.89%	16.29%
Diarrhoea	[redacted]	[redacted]	6.25%	0.00%	5.88%

Dyspnoea	[redacted]	[redacted]	8.33%	0.00%	6.67%
Fatigue	[redacted]	[redacted]	10.00%	2.78%	10.00%
Leukopenia	[redacted]	[redacted]	55.00%	0.00%	54.84%
Nausea	[redacted]	[redacted]	4.95%	2.78%	4.71%
Neutropenia	[redacted]	[redacted]	45.07%	8.33%	43.56%
Pyrexia	[redacted]	[redacted]	0.00%	2.78%	0.00%
Thrombocytopenia	[redacted]	[redacted]	37.60	19.44%	37.13%
Vomiting	[redacted]	[redacted]	2.65%	0.00%	3.08%
Source: calculations performed by the ERG, based on adverse events incidence tables from cost effectiveness model provided by the company *For SoC AE calculation, assumption was made (Weighted average of chemotherapy, bendamustine and investigational agents. See model safety tab)					

ERG comment: The ERG identified an error in the calculation of SoC adverse events incidence.

AE incidence for SoC, based on the weighted average of chemotherapy (38.46%), bendamustine (18.46%) and investigational agents (43.08%), was incorrectly calculated. Although it was assumed that investigational agents did not have AEs and therefore do not influence the number of events, the proportion of patients that received investigational agents should be included in the calculation of the sample size (N). By not doing this, the company over-estimated the relative SoC AE incidence. This is likely a favourable assumption for pembrolizumab, but is unlikely to be influential.

5.2.8 Health-related quality of life

HRQoL was measured in KEYNOTE-087. More specifically, EQ-5D-3L data were collected at treatment cycles 1-5 (i.e. every three weeks) and every 12 weeks up to 30 days post treatment discontinuation or until disease progression. Consistent with the NICE reference case, the UK social tariff[42] was used to obtain health state utility values from the responses on the EQ-5D-3L. Although the SLR also identified two relevant HRQoL studies, HRQoL data from KEYNOTE-087 were preferred by the company. It was unclear whether this was because the HRQoL studies identified in the SLR were inconsistent with the NICE reference case[43] or did not report utilities stratified by overall response.[44]

The company calculated utility values (Table 5.11) stratified by overall response (i.e. separately for patients with CR, PR and SD). However, this post hoc utility calculation was based on observations from week 12 in the KEYNOTE-087 trial only (i.e. ignoring observations at other time points). These utility scores were multiplied by the response rates from KEYNOTE-087 and Cheah et al, (2016)[7] to obtain the progression free health state utility values for pembrolizumab ([redacted] and [redacted] for cohort 1 and 2) and SoC ([redacted]) respectively (Table 5.11).

Similarly, response rates from Lafferty et al,[21] an abstract retrospectively reporting on single centre experiences with alloSCT in patients with Hodgkin Lymphoma, were used to calculate the post-alloSCT utility. Combining these response rates with the 12 week utilities (stratified by response) from KEYNOTE-087 resulted in a post-alloSCT utility of 0.865. To account for the possibility of acute graft versus host disease after alloSCT, a disutility of 0.15[45] is applied to 61.5%[21] of the patients for the first

14 weeks post-alloSCT. This resulted in a post-alloSCT utility of 0.773 for the first 14 weeks which was assumed to increase to 0.865 afterwards.

Table 5.11: Utility scores for the progression free (treatment dependent) and post-alloSCT disease health states

	Utility (12 week observations only)	Pembrolizumab response rates (cohort 1)	Pembrolizumab response rates (cohort 2)	SoC response rates	Post-alloSCT response rates
	KEYNOTE-087	KEYNOTE-087	KEYNOTE-087	Cheah et al[7]	Lafferty et al[21]
Total N	[redacted]	[redacted]	[redacted]	[redacted]	10
CR	[redacted]	[redacted]	[redacted]	[redacted]	70.0%
PR	[redacted]	[redacted]	[redacted]	[redacted]	30.0%
SD	[redacted]	[redacted]	[redacted]	[redacted]	0.0%
Utility[a]	[redacted]	[redacted]	[redacted]	[redacted]	0.865

Source: Economic model submitted by the company and CS Table 75
CR = complete response; PR = partial response; SD = stable disease;
[a]Utility was calculated by combining the Utility scores stratified by response and the response rates

The company did not use the PD utility score (of [redacted]) from KEYNOTE-087 arguing that this utility "is not predictive of a meaningful decrement in QoL", due to it being estimated based on 12 week observations only. Therefore, the company opted to use a utility decrement (of 0.33) calculated by subtracting the SD and PD utilities from Swinburn et al.[43] This resulted in a PD utility of [redacted].

Additionally, the company applied age related utility decrements, derived from UK population norms, in all health states (see CS Table 82). This was conditional on the starting age in the model (34 and 40 years for cohort 1 and cohort 2 respectively).

Finally, the company considered the impact of grade 3+ adverse events (see Section 5.2.7) on HRQoL. Given the absence of disutilities in relapsed or refractory Hodgkin Lymphoma, disutilities were identified in oncology and myocardial infarction (see CS Table 77 for a summary of sources). In case multiple sources were available an average was calculated. The disutilities and adverse event durations from the various adverse events are reported in CS Table 78. Table 5.12 below provides an overview of the calculated disutilities and the assumed duration of the AE. Multiplying the duration, the disutility and the occurrence of adverse events (see section 5.2.7) resulted in one-off disutilities of [redacted], [redacted] for pembrolizumab (separately for cohort 1 and cohort 2) and 0.0080 for SoC. These one-off disutilities were incorporated in the first cycle of the model.

Table 5.12: Adverse event disutilities

Adverse event (CTCAE grade 3+)	Disutility (per year with adverse event)	Duration (days)	Disutility (per occurrence of adverse event)
Anaemia	-0.0900	14.8	-0.0036
Diarrhoea	-0.1392	5.5	-0.0021
Dyspnoea	-0.0481	12.7	-0.0017
Fatigue	-0.1502	25.5	-0.0105
Leukopenia	-0.1264	12.1	-0.0042
Nausea	-0.1517	11.0	-0.0046
Neutropenia	-0.1264	12.3	-0.0042
Pyrexia	-0.1100	12.3	-0.0037
Thrombocytopenia	-0.1080	15.9	-0.0047
Vomiting	-0.1395	5.3	-0.0020

Source: Economic model submitted by the company and CS Table 83[1]

ERG comment: The ERG notes the following issues regarding the HRQoL data used by the company: (a) HRQoL data used by the company is restricted to observations from week 12 only, (b) using a decrement for progressive disease that is not from KEYNOTE-087, (c) progression free utility benefit for pembrolizumab maintained for patients without alloSCT, (d) sources for post-alloSCT HRQoL, (e) HRQoL consequences of disease progression post-alloSCT are not (explicitly) incorporated, (f) one technical error and one inconsistency in the calculation of the HRQoL.

(a) The company restricted the HRQoL data, used in its base-case, to KEYNOTE-087 observations from week 12 only. In response to clarification question B15, the company provided the results of mixed effects model analyses incorporating all observed EQ-5D data from KEYNOTE-087 (Table 5.13). Unfortunately, no diagnostics or goodness of fit statistics were provided by the company. Nevertheless, to utilise all available KEYNOTE-087 data, the ERG prefers to use utility scores generated by this mixed effects model. It is, however, notable that the coefficient for "PR versus CR" is positive, i.e. indicating a higher utility for PR than for CR. This lacks face validity, hence, the ERG decided to set this coefficient (0.01453) to zero. This resulted in a utility of [redacted] for both CR and PR while the estimated utility value for SD is [redacted]. Combining this with the observed response status, this resulted in PF utility values of [redacted] for pembrolizumab (cohorts 1 and 2) [redacted] for SoC. This SoC PF utility (of [redacted]) is more consistent with the SoC PF utility of 0.76 reported in TA462[6] than the SoC PF utility used in the CS base-case ([redacted]). Additionally, the PD utility changed to [redacted] while the post-alloSCT utility changed to 0.725 for the first 14 weeks and to 0.818 for after the first 14 weeks.

Table 5.13: Utilities estimated from mixed effects model using all observed EQ-5D data from KEYNOTE-087

Covariates	Estimated effect	Standard error
Intercept (reference = CR)	[redacted]	[redacted]
PR versus CR	[redacted]	[redacted]
SD versus CR	[redacted]	[redacted]
PD versus CR	[redacted]	[redacted]

Source: response to clarification question B15
Note: not marked as CiC in the clarification response
CR = complete response; PR = partial response; SD = stable disease; PD = progressive disease

(b) The company does not use the estimated PD utility from KEYNOTE-087 in its base-case. This was justified by stating that KEYNOTE-087 only contains observations shortly after progression and which might not capture the long-term utility decrement due to progression. Therefore, the company estimates the PD utility using the SD utility from KEYNOTE-087 and a utility decrement of 0.33 from Swinburn et al.[43] The ERG was not convinced that using this utility decrement from Swinburn et al.[43], over the PD utility estimated from KEYNOTE-087 is appropriate, given the company provided no evidence indicating a long-term impact of progression consistent with this utility decrement (of 0.33 versus SD). Additionally, the ERG in TA462 criticised the utilities from Swinburn et al.[43] by stating *"we suggest that the results from Swinburn and colleagues are outliers and may not be realistic. The Swinburn study used TTO methodology using estimates from the general public and it may be that their perception of the disease is not consistent with EQ-5D valuation."* This quote also highlights that the utilities from Swinburn et al.[43] deviate from the NICE reference case (as it is not consistent with EQ-5D valuation). Therefore, consistent with the NICE reference, the ERG's approach in TA462 (which was ultimately accepted by the committee), for the ERG base-case is to use HRQoL data from the pivotal trial (KEYNOTE-087) to estimate the PD utility (estimated PD utility based on mixed effects model is [redacted]). This PD utility estimate is more in line with the PD utility, estimated based on CheckMate 205, that was preferred by the ERG and accepted by the committee in TA462.[6]

(c) Based on response status (i.e. proportion of patients with CR, PR and SD), treatment specific PFS utilities are estimated and used throughout the model time horizon for the PFS health state. However, it is inconsistent to use the response status, combined for both groups of patients that undergo alloSCT and those who do not, to estimate the utilities for patients that do not undergo alloSCT. Particularly given that the response status for patients that undergo alloSCT is likely better than for patients who do not undergo alloSCT. Therefore, the ERG recalculated the post-12 week PFS utilities based on the response status of patients who did not undergo alloSCT. This resulted in utility values of [redacted] for pembrolizumab (both cohorts) and [redacted] for SoC. Based on the mixed model these utilities would be lower for pembrolizumab ([redacted] and [redacted] for cohort 1 and 2) and SoC ([redacted]).

(d) The company uses a disutility from Kurosawa et al.[45] (applied to 61.5% of the patients) to account for the possibility of acute graft versus host disease after alloSCT. This disutility is applied to the post-alloSCT utility estimated based on the KEYNOTE-087 estimates. In clarification question B18, the ERG questions why a disutility only is obtained from Kurosawa et al.[45]. The company stated that it was believed to be inappropriate to also use the utility estimate from Kurosawa et al.[45] due to the differences between populations in Kurosawa et al.[45] and KEYNOTE-087. The ERG, however, believes it is inappropriate to use KEYNOTE-087 utility estimates, including only one post-alloSCT observation (response to clarification question B15b), to estimate post-alloSCT utility values. Although the ERG recognises the differences between populations (i.e. in Kurosawa et al.[45] and KEYNOTE-087), given that Kurosawa et al.[45] is the only identified study to provide post-alloSCT preference-based (e.g. EQ-5D) utility measures (confirmed by the company in response to clarification question B18), the ERG prefers to use Kurosawa et al.[45] to obtain post-alloSCT utility values in the model. This resulted in a post-alloSCT utility of 0.708 for the first 14 weeks which was assumed to increase to 0.800 afterwards (these values were 0.773 and 0.865 in the CS base-case).

(e) Due to the lack of a post-alloSCT progression health state (see also ERG critique in section 5.2.2), it is questionable whether the impact of progression on HRQoL post-alloSCT is captured. Therefore, the ERG performed a scenario analysis to explore the impact of this assumption.

(f) Finally, the ERG identified a technical error in the calculation of the AE disutility (in the model the AE duration is divided, to convert from day to year, by 365.25 twice instead of once) as well as an inconsistency in the proportion of responders used to calculate PF utility estimates (see difference in response between CS Table 80 and the Table provided in response to clarification question B5). In the ERG base-case, the technical error was corrected and the number reported in response to clarification question B5 (updated version of CS Table 62, see Table 5.5 of the ERG report) is used to estimate PF utilities.

Table 5.14 below provides an overview of the utilities used in the ERG base-case (combining all abovementioned adjustments).

Table 5.14: Utilities used in the CS and ERG base-case

Health state		CS base-case utility	ERG base-case utility[a]
PF first 12 weeks	pembrolizumab cohort 1	[redacted]	[redacted]
	pembrolizumab cohort 2	[redacted]	[redacted]
	SoC	[redacted]	[redacted]
PF after first 12 weeks (no alloSCT)[b]	pembrolizumab cohort 1	[redacted]	[redacted]
	pembrolizumab cohort 2	[redacted]	[redacted]
	SoC	[redacted]	[redacted]
PD	treatment independent	[redacted]	[redacted]

Health state		CS base-case utility	ERG base-case utility[a]
Post-alloSCT first 14 weeks	treatment independent	0.773	0.708
Post-alloSCT after first 14 weeks	treatment independent	0.865	0.800

PR = progression free; PD = progressive disease;
[a]Standard error calculated by multiplying the estimated utility by 0.1 (consistent with the company's approach)
[b]The estimated PF utilities after the first 12 weeks (no alloSCT) would be [redacted] for pembrolizumab (cohort 1 and 2) and SoC respectively, when using the MSD survey only to estimate the proportion of patients receiving alloSCT conditional on response status (see section 5.2.6.2). These values were used in the final ERG base-case.

5.2.9 Resources and costs

5.2.9.1 Drug acquisition and administration costs

The electronic market information tool (eMit)[37] was used to acquire drug acquisition cost of pembrolizumab and components of SoC. When these were unavailable, costs from the British National Formulary[46] were used. Administration costs were obtained from the NHS reference costs[47] (see Table 5.15).

Pembrolizumab

The list price of 200 mg pembrolizumab was £5,260 (derived from the cost of 2 x 100 mg vials at £2,630 each per patient). Through a Commercial Access Agreement (CAA), [redacted]. As established previously in TA357[48] and TA428 [49], the NHS reference cost code SB 12Z [47] was used as administration cost, thereby adding £236.19 per 21 day cycle.

Standard of care

Consisting of chemotherapies (38.5%, each of the 12 treatments accounting for 3.2%), treatment with investigational agents (43.1%) and bendamustine (18.5%), drug acquisition costs for SoC varied by treatment agent. Acquisition and administration costs of investigational agents were assumed to be £0, for other components of SoC these costs are described in Table 5.15. For dosages/m^2, the number of vials required per administration was calculated based on a BSA of 1.85m^2 (SD 0.024). Upon request the company clarified that CS Table 56[1] contained an incorrect BSA but that the correct number was used in the model. For each component of SoC, the model assumed vial wastage and calculated the vial combination resulting in the lowest possible price for the required dosage. The treatment costs of SoC per seven day cycle (see Table 5.15) consist of the treatment costs of all SoC components. These were calculated by combining the calculated drug acquisition cost per cycle with the administration costs, adjusting these costs to the seven day timeframe and the proportion of patients treated within the SoC arm.

Best supportive care

BSC consisted of several subsequent treatments that are described in Table 5.16, which were selected based on the approach taken in TA462.[6] Acquisition costs and administration costs combined with the proportions of patients treated with each component of BSC resulted in a one-off cost of £4,848.22. In line with assumptions made in TA462,[6] palliative care and clinical trial treatment were assumed to have no costs.

Table 5.15: Treatment costs

Regimen	Acquisition cost/cycle	Administration costs/cycle	Cycle length (days)	Maximum number of cycles	Proportion of treatment (%)
Pembrolizumab	[redacted]	£236.19 [a]	21	35	100
Standard of care					
ICE	£1,230.82	£711.23 [b]	14	3	3.2
IVE	£2,183.65	£1,039.33 [c]	21	3	3.2
MINE	£1,209.02	£1,039.33 [c]	28	2	3.2
IVOx	£1,132.46	£1,039.33 [c]	21	3	3.2
IGEV	£2,109.48	£1,367.43 [d]	21	4	3.2
GEM-P	£100.86	£711.23 [b]	28	3	3.2
GDP	£93.06	£383.13 [e]	21	2	3.2
GVD	£1,491.60	£711.23 [b]	21	2	3.2
ESHAP	£63.32	£1,367.43 [d]	28	4	3.2
ASHAP	£68.73	£1,367.43 [d]	28	3	3.2
DHAP	£76.39	£383.13 [e]	21	2	3.2
DHAOx	£89.69	£383.13 [e]	21	4	3.2
Bendamustine	£123.30	£383.13 [e]	28	6	18.5

Source: CS Table 91, Table 92[1]
[a] Deliver Simple Parenteral Chemotherapy at First Attendance
[b] Delivering complex chemotherapy at first attendance and delivering a subsequent complex chemotherapy element within the same cycle
[c] Delivering complex chemotherapy at first attendance and delivering a subsequent complex chemotherapy element within the same cycle
[d] Delivering complex chemotherapy at first attendance and delivering three subsequent complex chemotherapy elements within the same cycle
[e] Delivering complex chemotherapy at first attendance

Table 5.16: BSC

Therapy	Distribution of patients across therapies (%)	Cycle length	Number of cycles	Acquisition costs/cycle	Administration costs/cycle
Gemcitabine monotherapy (administered over 4 weeks)	8.33	28 days	4.0	£47.76	£236.19
RVIG	16.67	21 days	4.5	£3,299.29	£1,367.43
DHAP	11.67	21 days	6.0	£76.39	£383.13
CHOP	1.67	21 days	6.0	£32.45	£383.13
IVAC	3.33	21 days	3.5	£1,832.00	£1,695.53
Weekly therapy (PMitCEBO)	8.33	14 days	7.0	£109.11	£711.23
Palliative care	46.67				
Clinical trial treatment	3.33				

Source: CS Table 93, Table 94, Table 97,[1] Model

CHOP = cyclophosphamide, doxorubicin, prednisolone, vincristine; DHAP = dexamethasone, cytarabine, cisplatin; IVAC = cytrabine, etoposide, ifosfamide, mesna; PMitCEBO = bleomycin, cyclophosphamide, etoposide, mitoxantrone, prednisolone, vincristine; RVIG = gemcitabine, ifosfamide, mesna, prednisolone, rituximab, vinorelbine

ERG comments: The ERG identified the following inconsistencies and assumptions lacking justification: (a) potential over-estimation of SoC costs due to the assumed mix of chemotherapy regimens within SoC, (b) the lack of missed doses, and (c) the number of cycles used for the components of BSC.

(a) The assumption that all chemotherapy agents contribute equally to the mix of SoC is not justified by the company. Responding to clarification question B19 [10], the company explains that *"There is a paucity of evidence on the preferred or standard mix of chemotherapy regimens given to patients in UK clinical practice"*, and an approach previously accepted was used. Given the extensive efforts taken by the company to interview clinical experts on alloSCT uptake, it can be questioned why the comparator treatment mix was not a topic discussed with the clinical experts. In TA462 it is stated that *"Clinical advice to the ERG suggests that gemcitabine regimens such as GDP (gemcitabine, dexamethasone, cisplatin) are commonly used in this patient population in the UK but platinum-containing regimens such as ESHAP (etoposide, methylprednisolone, cytarabine, cisplatin) and DHAP (dexamethasone, cytarabine, cisplatin) are also in common use.*"[6] Given the regimens mentioned are of lower price than other chemotherapy regimens, the ERG wishes to point out that the company has likely overestimated costs of SoC, an assumption that favours pembrolizumab.

(b) The company states that due to a lack of information on missed doses in the SoC arm, missed doses were not incorporated for pembrolizumab or SoC. The ERG wishes to highlight that the incremental cost of pembrolizumab versus SoC could be biased, even though the effect is likely to be small.

(c) The company assumed treatment durations for components of BSC to be shorter than the maximum number of cycles. The ERG recognises this assumption to be conservative and possibly reflective of treatment intensity after repeated progression. However, the assumption and treatment durations used lack justification.

5.2.9.2 Health-state costs

Lacking detailed published data on resource use in the patient population, data used to inform health-state costs stemmed from previous TAs. In both non-alloSCT health-states (pre- and post-12 weeks), i.e. progression-free and progressed disease, monitoring costs consisted of outpatient attendance, blood tests and CT and PET scans (see Table 5.17) and amounted to £68.78 per week.

Table 5.17: Weekly monitoring costs

Unit	Unit cost	Usage per week	Weekly cost
Outpatient attendance	£173.17	0.20	£34.52
Blood count	£3.10	0.20	£0.62
Biochemistry	£1.18	0.20	£0.24
CT scan	£120.99	0.06	£6.96
PET scan	£920.24	0.03	£26.45
Source: Model			

Progressed disease
Upon disease progression, initial treatment with pembrolizumab or SoC is discontinued and costs for BSC are applied as a one-off event.

Alive post-alloSCT
Post-alloSCT health-state costs were taken from Radford [29], a study reporting on costs in 14 relapsed or refractory classical Hodgkin Lymphoma patients treated with alloSCT, the source for resources and costs preferred by the committee of TA462 [6]. These costs were applied once upon treatment with alloSCT and were assumed to consist of alloSCT treatment costs, monitoring costs, costs of adverse events, costs of subsequent treatment and terminal care costs. No long-term costs were added.

Adverse event costs
A selection of grade 3+ AEs costs, based on previous appraisal TA462[6] and validated in a clinician survey, was applied dependent on treatment. Assuming that serious AEs lead to the discontinuation of treatment, patients on treatment beyond the first year were assumed to be free from AEs, and investigational agents were assumed to have no AEs. Resource use and costs of AEs were taken from the NHS reference costs by means of a weighted average of HRG codes, applied to the model as one-off event costs of [redacted], [redacted] for pembrolizumab in cohort 1 and cohort 2 respectively, and £1,945.74 for SoC in both cohorts.

Table 5.18: Adverse event costs

Adverse Event	Unit Cost	Source	Rational
Anaemia	£814.03	NHS reference costs 2015-16 [47]	TA411 TA399 TA391
Diarrhoea	£1,497.86	NHS reference costs 2015-16 [47]	TA391 TA440
Dyspnoea	£718.76	NHS reference costs 2015-16 [47]	TA420
Fatigue	£1,499.09	Brown (2013) [50] and NHS reference costs 2011-12[51] inflated with HCHS index	TA391
Leukopenia	£1,142.90	NHS reference costs 2015-16 [47]	TA391
Nausea	£872.42	NHS reference costs 2015-16 [47]	TA411
Neutropenia	£1,142.90	NHS reference costs 2015-16 [47]	TA411 TA399
Pyrexia	£3,923.50	NHS reference costs 2013-14 [52] inflated with HCHS index	TA366 TA311
Thrombocytopenia	£636.19	NHS reference costs 2015-16 [47]	TA399 TA440
Vomiting	£1,497.86	NHS reference costs 2015-16 [47]	TA360 TA440
Source: CS Table 99 [1]			

Terminal care costs
Terminal care costs are applied at death of patients in the non-alloSCT health states to reflect increased health care consumption in the period before death. The proportions of patients treated in different settings were taken from a population of non-small cell lung cancer patients (see Table 98 of the CS [1]).

Cost of terminal care resources stemmed from the same source but were updated with 2015-2016 NHS reference costs or increased for inflation with the HCHS hospital and community health service index. Hospital care, hospice care and homecare consisting of GP visits, nurse visits and drugs amounted to a total of £4,064.64 terminal care costs.

ERG comment: The ERG considers the costs associated with alloSCT to be under-estimated. In the model, a one-off cost was applied upon treatment with alloSCT. The company argues that it includes costs and resource use of alloSCT treatment, monitoring costs, costs of adverse events, of subsequent treatment and terminal care costs. The ERG wishes to point out that the company deviated from the methods in TA462 where the one-off cost was used in a scenario analysis, however, monthly costs for subsequent treatment and monitoring were added that were foregone in this TA. In their response to clarification question B22.a,[10] the company did not specify how the one-off cost based on a mean follow-up period of 3.44 years and an unknown proportion of deaths could reflect costs of a lifetime horizon. The ERG therefore applied monitoring costs, comparable to those used in TA462, over the lifetime horizon in their base-case, showing that the company's analysis favoured pembrolizumab.

5.2.10 Cost effectiveness results

In the deterministic base-case analysis, total QALYs and LYs gained were larger in the pembrolizumab treatment arm compared to UK SoC in both cohorts. Tables 5.19 and 5.20 show that the main benefit of pembrolizumab versus SoC are mostly due to QALY gains beyond week 12 with alloSCT (71% and 78% of incremental QALYs in cohort 1 and cohort 2 respectively). Total costs were also higher for pembrolizumab than for SoC. Incremental costs mainly resulted from differences in acquisition costs and alloSCT costs between pembrolizumab and SoC. Pembrolizumab treatment resulted in deterministic incremental cost effectiveness ratios (ICERs) of £43,511 and £48,571 per QALY gained for cohort 1 and cohort 2 respectively, as per the company's corrected base-case (Table 5.21).

Table 5.19: Cohort 1 QALYs breakdown (discounted)

	Week 0 to week 12		**Beyond week 12 (without alloSCT)**		**Beyond week 12 (with alloSCT)**
Base-case					
	PF	**PD**	**PF**	**PD**	**Alive**
Pembrolizumab	0.186	0.001	0.684	0.664	2.861
SoC	0.166	0.011	0.107	0.951	1.522
Corrected base-case					
Pembrolizumab	0.186	0.001	0.655	0.638	3.016
SoC	0.166	0.011	0.089	0.845	2.112
Source: (corrected) cost effectiveness model submitted by the company					

Table 5.20: Cohort 2 QALYs breakdown (discounted)

	Week 0 to week 12		**Beyond week 12 (without alloSCT)**		**Beyond week 12 (with alloSCT)**
Base-case					
	PF	**PD**	**PF**	**PD**	**Alive**
Pembrolizumab	0.186	0.001	0.457	0.745	2.503
SoC	0.180	0.003	0.113	0.960	1.455

Corrected base-case					
Pembrolizumab	0.186	0.001	0.426	0.701	2.757
SoC	0.180	0.003	0.092	0.849	2.076
Source: (corrected) cost effectiveness model submitted by the company					

Table 5.21: Company base-case results

Technologies	Cohort	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental QALYs	ICER (£) versus baseline (QALYs)
Base-case							
UK SoC	Cohort 1	44,278	4.385	2.757	-	-	-
	Cohort 2	43,275	4.330	2.711	-	-	-
Pembrolizumab	Cohort 1	106,908	6.153	4.397	62,630	1.639	38,201
	Cohort 2	92,100	5.594	3.892	48,825	1.181	41,341
Corrected base-case							
UK SoC	Cohort 1	52,017	4.864	3.223	-	-	-
	Cohort 2	51,424	4.832	3.200	-	-	-
Pembrolizumab	Cohort 1	107,459	6.252	4.497	55,442	1.274	43,511
	Cohort 2	93,732	5.775	4.072	42,308	0.871	48,571
Sources: CS Table 101[1], corrected cost-effectiveness model submitted by the company ICER = incremental cost-effectiveness ratio; LYG = life years gained; QALYs = quality-adjusted life years							

ERG comment: The ERG's concern relates to the exclusion of BSC as a comparator in the base-case analysis. BSC was not included as a comparator in the base-case, and therefore pembrolizumab could not be compared to all relevant alternatives at the same time.

5.2.11 Sensitivity analyses

The company performed and presented probabilistic sensitivity analysis (PSA) and deterministic sensitivity analysis (DSA) in order to quantify the uncertainty surrounding the company's results.

Compared with the deterministic results, the PSA with 1,000 iterations showed a comparable relative decrease in incremental costs and QALYs, which did not result in large changes to the ICER of cohort 1 (£43,653). In cohort 2, the PSA showed decreased incremental costs and even larger (relative) decreased incremental QALYs compared with the deterministic results, which resulted in an ICER of £50,894 (Table 5.22).

Cost effectiveness acceptability curves (CEACs) showed that there was a 60.1% (cohort 1) and 50.4% (cohort 2) probability of pembrolizumab to be cost effective compared to SoC at a willingness to pay (WTP) of £50,000 per QALY (Figures 5.5 and 5.6). However, these probabilities are reduced to 1.1% and 1.4% respectively at a WTP of £20,000 per QALY, and 20.5% and 16.1% at a WTP of £30,000 per QALY.

The company stated that DSAs were conducted for all key variables. Parameters were varied within their 5% and 95% confidence intervals where possible, and +/- 10% otherwise. The DSA results were presented in tornado diagrams including the 15 key model drivers. The following parameters were identified as most influential on the cost effectiveness of pembrolizumab versus SoC:

Cohort 1:

1. Discount rate – Outcomes (0.035; 0.000-0.060)

2. Response at week 12 – SoC – CR odds ratio ([redacted])
3. Response at week 12 – SoC – PR odds ratio ([redacted])

Cohort 2:

1. Response at week 12 – SoC – CR odds ratio ([redacted])
2. Discount rate – Outcomes (0.035; 0.000-0.060)
3. Response at week 12 – SoC – PR odds ratio ([redacted])

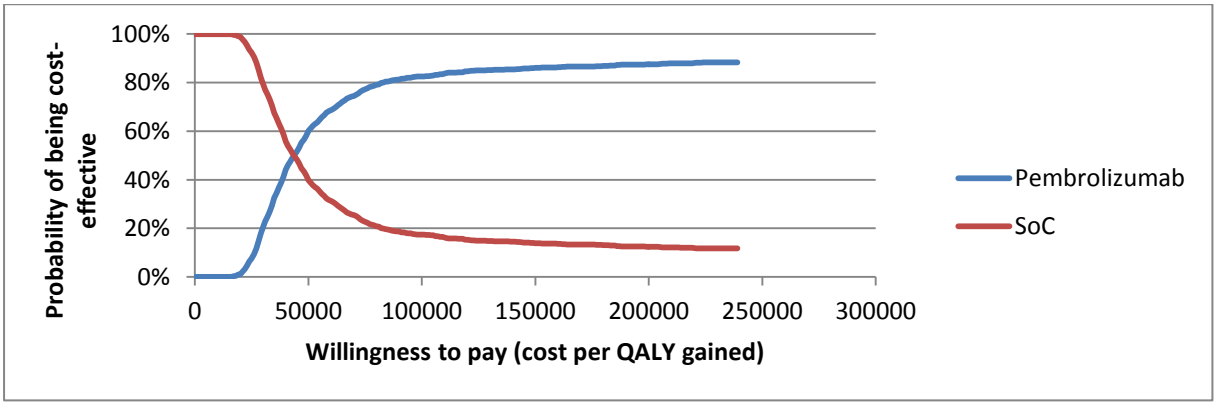
The WTP threshold of £50,000 per QALY was exceeded in the outcomes discount rate parameter and the CR odds ratio of SoC at week 12 response for cohort 1. For cohort 2, the WTP threshold was exceeded in all of the three abovementioned parameters.

Table 5.22: Incremental cost effectiveness results based on PSA (discounted, with CAA, 1,000 simulations)

Technologies	Cohort	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	ICER (£) versus baseline (QALYs)
Base-case						
UK SoC	Cohort 1	46,723	2.857	-	-	-
	Cohort 2	45,391	2.771	-	-	-
Pembrolizumab	Cohort 1	106,672	4.361	59,949	1.505	39,841
	Cohort 2	92,941	3.875	47,550	1.105	43,049
Corrected base-case						
UK SoC	Cohort 1	53,491	3.219	-	-	-
	Cohort 2	54,028	3.254	-	-	-
Pembrolizumab	Cohort 1	106,702	4.438	53,211	1.219	43,653
	Cohort 2	94,522	4.050	40,494	0.796	50,894

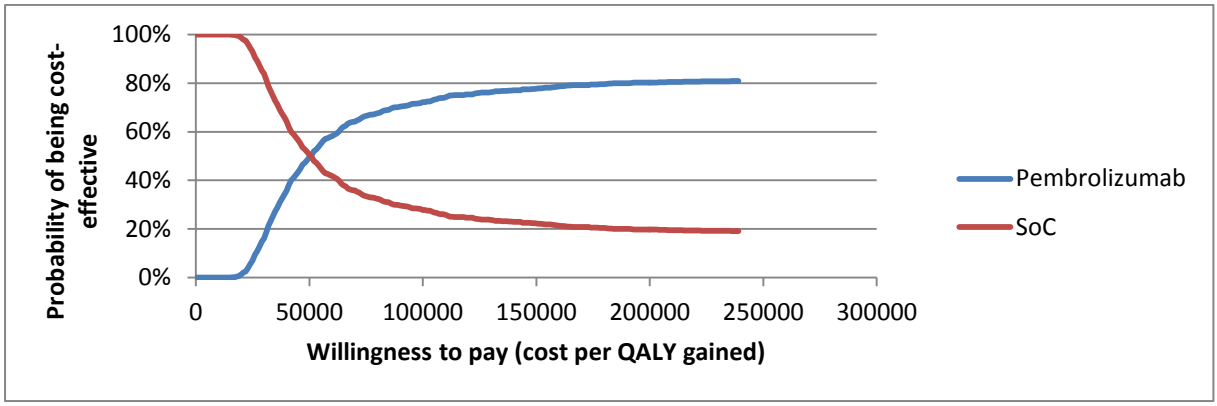
Sources: CS table 101, cost-effectiveness model after correction by company
ICER = incremental cost-effectiveness ratio; LYG = life years gained; QALYs = quality-adjusted life years

Figure 5.5: Cohort 1 cost effectiveness acceptability curve (discounted, with CAA)



Source: corrected cost effectiveness model provided by the company.

Figure 5.6: Cohort 2 cost effectiveness acceptability curve (discounted, with CAA)



Source: corrected cost effectiveness model provided by the company.

The following five scenario analyses were performed by the company (Table 5.23). The results shown are based on the company's corrected base-case.

Scenario 1: assessing BSC as a comparator as per the NICE scope

Scenario 2: assessing different alloSCT rates

a. 100% alloSCT in patients with CR, PR or SD
b. Alternative lower PR alloSCT rate from MSD clinician survey

Scenario 3: using MAIC HR and OR rather than naïve ITC

Scenario 4: Alternative extrapolation scenarios to estimate PFS and OS

a. Considering a Weibull curve for week 0-12 PFS extrapolation in cohort 2
b. Considering a Gompertz curve for week 12+ PFS extrapolation in cohort 2
c. Considering a Lognormal curve following alloSCT

Scenario 5: assessing varying the time horizon to 50 years

Across all the scenarios, the ICER ranged between £23,564 and £47,957 for cohort 1, and between £24,492 and £56,677 for cohort 2. Scenario 2a had the biggest impact on the ICER in both cohorts (ICER decrease of approximately £20,000 and £24,000 for cohort 1 and 2 respectively).

Table 5.23: Results from the scenario analyses based on the company's corrected base-case

Scenario	Cohort	Pembrolizumab			UK SOC			Pembrolizumab vs UK SOC		
		Total costs	Total LYs	Total QALYs	Total costs	Total LYs	Total QALYs	Inc. costs	Inc. QALYs	ICER
Company's corrected base case	Cohort 1	£107,459	6.252	4.497	£52,017	4.864	3.223	£55,442	1.274	£43,511
	Cohort 2	£93,732	5.775	4.072	£51,424	4.832	3.200	£42,308	0.871	£48,571
Scenario 1	Cohort 1	£107,459	6.252	4.497	£51,188	4.864	3.223	£56,270	1.274	£44,161
	Cohort 2	£93,732	4.832	3.200	£50,713	4.832	3.200	£43,018	0.871	£49,387
Scenario 2a	Cohort 1	£119,943	8.503	6.768	89,436	7.175	5.474	£30,507	1.295	£23,564
	Cohort 2	£116,185	8.261	6.537	£87,472	7.053	5.364	£28,713	1.172	£24,492
Scenario 2b	Cohort 1	£106,221	6.029	4.272	£49,951	4.736	3.098	£56,270	1.173	£47,957
	Cohort 2	£91,431	5.520	3.819	£49,360	4.705	3.077	£42,070	0.742	£56,677
Scenario 3	Cohort 1	£107,459	6.252	4.497	£45,292	4.419	2.790	£62,166	1.707	£36,423
	Cohort 2	£93,732	5.775	4.072	£46,944	4.558	2.933	£46,787	1.139	£41,087
Scenario 4a	Cohort 2	£93,261	5.766	4.062	£51,234	5.814	3.175	£42,027	0.886	£47,410
Scenario 4b	Cohort 2	£93,439	5.688	4.000	£51,500	4.852	3.217	£41,938	0.783	£52,562
Scenario 4c	Cohort 1	£107,459	6.451	4.642	£52,016	5.003	3.324	£55,442	1.318	£42,075
	Cohort 2	£93,732	5.957	4.204	£51,423	4.969	3.300	£42,308	0.904	£46,812
Scenario 5	Cohort 1	£107,459	6.377	4.582	£52,016	4.951	3.283	£55,442	1.300	£42,651
	Cohort 2	£93,732	5.889	4.150	£51,423	4.918	3.259	£42,308	0.890	£47,516
Source: CS Table 20.[53]										

ERG comment: The ERG had minor concerns about (a) the choice of variation in the DSA, (b) the cost effectiveness probability of pembrolizumab at lower WTP thresholds in the PSA, (c) inappropriate parameters in the PSA, and (d) an insufficient number of iterations in the PSA.

(a) In variables for which it was not possible to use 5% and 95% confidence intervals, a variation of +/- 10% was chosen without providing any rationale for this decision. Additionally, the ERG believes this variation may be small when wishing to assess the full impact on the ICER.

(b) The probability of pembrolizumab being cost effective at WTP thresholds of £20,000 and £30,000 is much lower compared to the base-case WTP, indicating the CEAC gradient to be very steep.

(c) Patient characteristics (proportion female, average weight, body surface area) were included in the PSA, although they are considered first order uncertainty and typically not reflected in cohort model PSAs.

(d) The company ran the PSA on 1,000 iterations. The ERG concluded that this number was insufficient to test the robustness of the model, and therefore re-ran the analysis on 10,000 iterations

5.2.12 Model validation and face validity check

5.2.12.1 Face validity

The selected time-to-event models and health state utility values for the base-case analysis were validated by UK clinical experts. No detail was provided in the CS concerning the expert elicitation method and the number of experts consulted.

AlloSCT rates, which were obtained by using UK clinical expert opinion through a survey performed by the company and joining these with survey results from an existing survey, have been compared to alloSCT rates reported in previous studies. The rates used by the company were higher than in a French study[54] and lower than the rates reported in Cheah et al. (2016),[7] which were considered too low and too high in TA462, respectively. Several responses from the survey conducted by the company indicated that alloSCT could be administered after PD. However, this assumption was not included in the model following further discussions with UK clinicians because it was not thought to be UK standard practice. Additionally, the alloSCT rates have been validated by a UK clinical expert in this area. This expert suggested that alloSCT rates would be higher than the ones used in the cost effectiveness model, with alloSCT rates in the PR as high as in CR.

5.2.12.2 Internal validity

The company submission states that: "*the model structure, assumptions and rationale were critically reviewed by an independent health economics modelling expert*."[1, 39]

5.2.12.3 External validity

The survival estimates obtained from the cost effectiveness model were validated against the studies used to inform PFS and OS estimates of the model. The outcomes obtained for pembrolizumab and SoC from the cost effectiveness model were compared to the KEYNOTE-087 trial (Table 5.24). From this comparison, the company concluded that the proportions of patients in pre-progression and surviving at different points in time were similar in the model and KEYNOTE-087 and the SoC sources.[1, 7, 21]

5.2.12.4 Cross validity

No cross-validation of the model assumptions, model structure and model outcomes were performed with the previous TA462 in the same indication.[6]

Table 5.24: Comparison of model and trial outcomes

		Pembrolizumab			UK SoC		
Outcome		**Base case**	**KEYNOTE -087**	**ERG retrieval from the model**	**Base case**	**Cheah et al[7]**	**ERG retrieval from the model**
% PFS at 1 Year *	Cohort 1	54.79%	[redacted]	59.44%	4.1%[a]	~7.5%[a] -	3.97%
	Cohort 2	39.07%	[redacted]	43.75%	4.9%[a]		4.77%
OS at week 12	Cohort 1	98.96%	[redacted]	98.96%	98.96%	~100%	98.96%
	Cohort 2	98.76%	[redacted]	98.78%	98.76%		98.78%
OS at 72 Months**	Cohort 1	28.00%	-	15.50%	15.00%	15.00%	10.87%
	Cohort 2	22.00%		12.76%	15.00%		10.95%
OS after alloSCT 5 years		Base case	KEYNOTE-087		Base case	Lafferty et al.[21]	
	Cohort 1	54.50%	-	51.22%	54.50%	53.47%	51.22%
	Cohort 2						

Source: adapted from CS Table 102[1]
alloSCT = allogeneic stem cell transplantation; OS = overall survival; PFS = progression-free survival; SoC = standard of care
*using data post week 12 assuming no alloSCT as per KEYNOTE-087 design
** when no alloSCT is assumed as per assumption made about Cheah SoC arm
[a] Provided in the response to the clarification letter[10]

ERG comment: The main ERG concerns about model validation are (a) the non-reproducibility of Table 5.24, (b) the proportion of patients in the stable disease response status at 12 weeks, (c) the lack of cross-validation with TA462.

(a) The ERG attempted to retrieve the model outcomes presented in Table 5.24 but consistently retrieved different figures than provided in the CS. Additionally, this table (based on Table 102 of the CS) reports five-year OS after alloSCT from Lafferty et al.[21], which is derived from KM estimates made available in TA462. These should be interpreted with extreme caution, because the plateau at the end of these KM estimates (starting at approximately 20 months) may be caused by censoring. The abstract only reported one-year OS after alloSCT (69%). After the large number of events in the first year, it would be implausible for the rate of events to slow down that considerably. The ERG is concerned by the validity of the figures provided in Table 5.24 and considers Table 5.24 to be potentially misleading.

(b) The company assumed that all patients, who did not completely or partially respond and who did not progress or die at the 12-week decision nodes, were in the stable disease response category. Patients with a non-evaluable response are consequently automatically included in the stable disease response category. This assumption probably leads to an overestimation of the proportion of patients in the stable disease response status compared to KEYNOTE-087.

(c) Complete cross validation with TA462 was not performed by the company in both the CS and clarification response. The main differences between TA462 and the current assessment are the model structure, and how alloSCT is incorporated in the cost effectiveness model. TA462 used a three health states (progression-free, progressed, dead) semi-Markov model while the current model is composed of a short-term component (first 12 weeks), a decision tree element (at 12 weeks) and a long-term component (after 12 weeks). Additionally, progression was not allowed post-alloSCT in the current assessment while it was incorporated in TA462. Different assumptions were also made concerning the composition of SoC between the two assessments. All these discrepancies may have influenced the health benefits and costs obtained in the SoC arm. Table 5.25 compares the results of SoC between TA462 and the current assessment. The health benefits obtained from SoC were almost doubled and the costs of SoC were more than doubled in the current assessment compared to TA462. These discrepancies are most likely explained by the fact that patients in TA462 may receive alloSCT after 6 months while patients are considered for alloSCT after 12 weeks in the current assessment. These different assumptions have likely influenced health benefits and costs of SoC.

Table 5.25: Comparison of SoC results between TA462 and the current assessment

Assessment	Total QALY	Total costs
TA462[a]	1.870	£23,668
Current assessment[b]	3.684	£52,017

[a] Outcomes considered as the AC's most plausible analysis, retrieved from the committee papers for the second AC meeting, Table 4 of the ERG commentary on the company additional evidence
[b] Retrieved from the corrected company's cost effectiveness model, post clarification response, Cohort 1

5.3 Exploratory and sensitivity analyses undertaken by the ERG

Table 5.26 summarises all main issues highlighted by the ERG in Section 5.2, indicates the expected direction of bias introduced by these issues and whether these are examined in any analyses/incorporated in the ERG base-case.

Table 5.26: Main ERG critique of company's submitted economic evaluation

Issue	Bias introduced[a]	ERG analyses	Addressed in company analysis?
Model structure (section 5.2.2)			
• Incorporation of alloSCT at 12 weeks only	+/-	None	Not addressed
• No lag between decision and procedure	+	None	Not addressed
• No progressed disease health state post-alloSCT	+	SA	Not addressed
Population, interventions and comparators, perspective and time horizon (sections 5.2.3-5)			
• Comparator data based on mix of cohort 1 and 2	+ cohort 1, - cohort 2	None	Not addressed
• BSC only in scenario analysis	+/-	None	Not addressed
• Time horizon of 40 years	-	BC (FV)	Addressed in SA
Treatment effectiveness and extrapolation (section 5.2.6)			
• Use of alternative sources due to immature OS data from KEYNOTE-087	+/-	None	Requested, partially addressed
• Single-arm study used to inform treatment effectiveness	+/-	None	Not addressed

Issue	Bias introduced[a]	ERG analyses	Addressed in company analysis?
• Use of naive indirect treatment comparison	-	SA	MAIC explored in SA
• Over-estimation of post-alloSCT OS based on Lafferty et al	+	SA	Not addressed
• Curves derived from entire study data fitted to pre-12 week period	+/-	None	Not addressed
• Inflated SD health state due to patients with non-evaluable response status being considered to have SD	+/-	None	Not addressed
• Combining of MSD and BMS surveys likely introduces bias	+	BC (FV)	Addressed in SA
• Patients with PD cannot receive alloSCT	+	BC (FV)	Addressed in SA
• Increased uncertainty in post-12 week PFS due to fitting curves from 12 weeks onwards	+	SA	Partially addressed in SA
• HRs equal for pre- and post-12 week periods	+/-	None	Requested, explored in SA
• PFS used as proxy for TTD pre-12 weeks without justification	-	None	Not addressed
• TTD capped at 24 months	+	SA	Addressed in SA
Health-related quality of life (section 5.2.8)			
• Utilities only derived from 12-week observations	+	BC (MJ)	Company provided mixed model utilities
• Progressed disease utility not from KEYNOTE-087	+	BC (MJ)	Not addressed
• PFS utility for patients without alloSCT calculated based on patients with and without alloSCT	+	BC (MJ)	Not addressed
• Not using post alloSCT utilities from Kurosawa et al.[45] (only disutilities are used)	+	BC (MJ)	Not addressed
• Inconsistency with treatment effectiveness section regarding calculation of proportion of responders	+/-	BC (MJ)	Not addressed
Resources and costs (section 5.2.9)			
• Likely over-estimation of SoC resource use and costs due to SoC chemotherapy mix	+	None	Not addressed
• Under-estimation of post-alloSCT costs	+	BC (FV)	Requested, not addressed
• Missed doses not incorporated	+/-	None	Not addressed
Cost-effectiveness analyses (sections 5.2.10 and 5.2.11)			
• Exclusion of BSC from base-case	+/-	None	Requested, not addressed
• Patient characteristics included in PSA	+/-	BC (FE)	Not addressed
Validation (section 5.2.12)			
• Complete cross validation with TA462 not performed	NA	None	Not addressed

Issue	Bias introduced[a]	ERG analyses	Addressed in company analysis?
BC = base-case; FE = fixing error; FV = fixing violations; MJ = matters of judgement; NA = not applicable; SA = scenario analysis [a]Likely conservative assumptions (of the intervention versus all comparators) are indicated by '-'; while '+/-' indicates that the bias introduced by the issue is unclear to the ERG and '+' indicates that the ERG believes this issue likely induces bias in favour of the intervention versus at least one comparator.			

Based on all considerations from Section 5.2 (summarised in Table 5.26), the ERG defined a new base-case. This base-case included multiple adjustments to the original base-case presented in the previous sections. These adjustments made by the ERG form the ERG base-case and were subdivided into three categories (derived from Kaltenthaler 2016):[55]

- Fixing errors (correcting the model where the company's submitted model was unequivocally wrong)
- Fixing violations (correcting the model where the ERG considered that the NICE reference case, scope or best practice had not been adhered to)
- Matters of judgement (amending the model where the ERG considers that reasonable alternative assumptions are preferred)

Additionally, exploratory sensitivity analyses were performed by the ERG to examine the potential impact of alternative assumptions on the cost effectiveness estimates.

The ERG's base-case:

Fixing errors

1. Error in the calculation of AE disutilities
   The ERG corrected the error.
2. Patient characteristics included in the PSA
   The ERG corrected this by excluding patient characteristics from the PSA.

Fixing violations

3. Combining MSD and BMS surveys for obtaining the probabilities of alloSCT uptake conditional on response.
   The ERG used the MSD survey only.
4. Time horizon of 40 years, despite some patients still being alive at that point.
   The ERG used a time horizon of 50 years.
5. Model excludes long-term monitoring costs post-alloSCT.
   The ERG included these consistent with committee's preferences in TA462.

Matters of judgment

6. Use of utility values estimated based on observations from week 12 only; progressed disease utility was estimated based on an alternative source; PFS utility for patients without alloSCT calculated based on patients with and without alloSCT; not using post alloSCT utilities from Kurosawa et al.[45] (only disutilities); and inconsistency with treatment effectiveness section regarding calculation of proportion of responders
   The ERG used the mixed model utilities provided by the company and the literature (Kurosawa et al.[45]) to calculate alternative utilities (see section 5.2.8 for more details).
7. Distributions for pre-12 weeks OS over-estimates mortality.
   The ERG used alternative distributions (exponential for cohort 1, lognormal for cohort 2) for pre-12 weeks OS.

8. Proportion of patients in PD state receiving alloSCT was set to 0.
   The ERG used the results from MSD's clinician survey to inform this.

5.3.1 Probabilistic ERG base-case

The ERG performed a PSA to obtain the ERG base-case incorporating all abovementioned adjustments. This resulted in probabilistic ICERs of £64,186 and £78,696 per QALY gained for pembrolizumab (with CAA) versus SoC for cohorts 1 and 2 respectively (Table 5.27). The individual effects of each change on costs, QALYs and ICERs are presented in Section 6, Table 6.1. For comparison, the deterministic ERG base-case ICERs were £61,705 and £73,594 per QALY gained, for cohorts 1 and 2 respectively.

Table 5.27: ERG base-case (probabilistic)

	Technologies	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	Pembrolizumab ICER (£/QALY)
Cohort 1	Pembrolizumab	£108,894	4.602			
	SoC	£53,729	3.743	£55,165	0.859	£64,186
Cohort 2	Pembrolizumab	£93,953	4.277			
	SoC	£53,487	3.763	£40,466	0.514	£78,696

ERG = Evidence Review Group; ICER = incremental cost effectiveness ratio; QALY = quality-adjusted life year

The CEACs based on the ERG base-case (Figures 5.7 and 5.8) show that pembrolizumab has a probability of being cost effective of 18% and 42% for cohort 1 and 21% and 40% for cohort 2 at thresholds of £30,000 and £50,000 per QALY gained, respectively.

Figure 5.7: Cost effectiveness acceptability curve for ERG base-case (cohort 1)

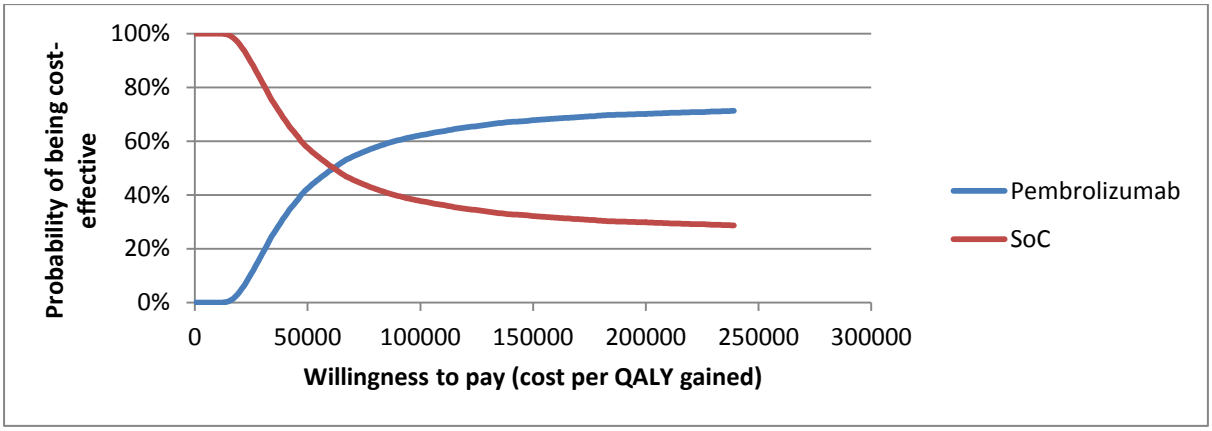
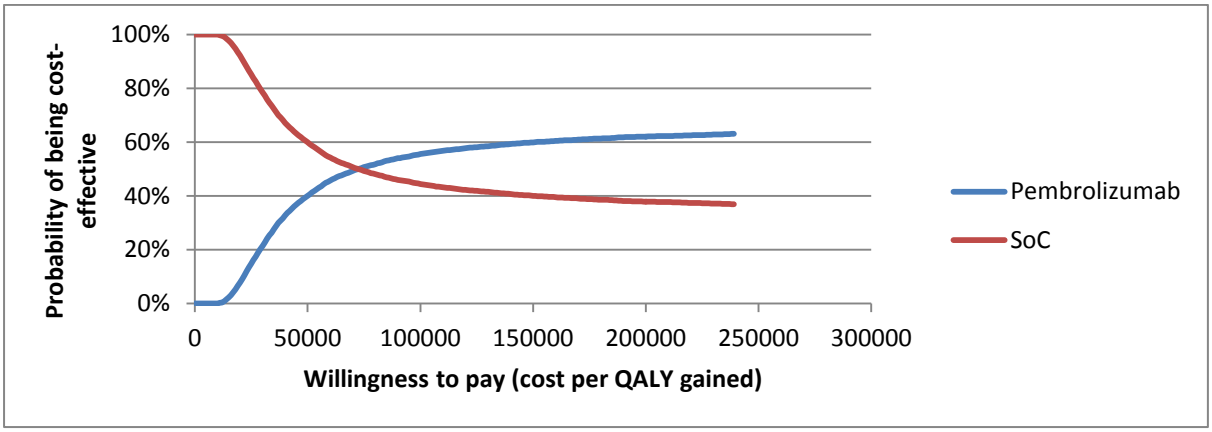


Figure 5.8: Cost effectiveness acceptability curve for ERG base-case (cohort 2)



5.3.2 Additional exploratory analyses performed based on the ERG base-case

Additional sensitivity analyses were performed to examine the potential impact of the following alternative assumptions on the cost effectiveness estimates. These were all performed using the ERG base-case. Results are presented in Table 6.2 in Section 6.

Exploratory analyses using the ERG base-case:

1. Alternative parametric survival models:
   Cohort 1: a) for post-12 weeks PFS (Gompertz)
   Cohort 2: a) post-12 weeks PFS (Gompertz) and b) post-12 weeks PFS (generalised gamma)
2. Use of MAIC instead of the naive indirect treatment comparison for estimating PFS hazard ratios and response rates at 12 weeks
3. Remove the 24-months cap on TTD
4. Use lower post-alloSCT utility (i.e. the PD utility) to explore the impact of ignoring PD after alloSCT
5. Use of alternative assumptions to extrapolate post-alloSCT OS from Lafferty et al (2017)

5.3.3 Subgroup analyses performed based on the ERG base-case

No subgroup analyses were performed.

5.4 Conclusions of the cost effectiveness section

Reviewing the overall evidence, the ERG confirmed that there was no existing cost effectiveness model for pembrolizumab for the current indication, and thus that development of a de novo model was necessary. The economic model described in the CS is considered by the ERG to meet the NICE reference case, with the exceptions of (1) the exclusion of a comparator that was identified in the scope, and (2) a slightly short time horizon. The absence of BSC from the main analysis was justified by a lack of data, and has been accepted by the committee in previous appraisal TA462. Another potential comparator in the future may be nivolumab, which has recently been recommended for use in part of the present population (cohort 1). The time horizon was extended (from 40 to 50 year) by the company to cover patients' lifetime in scenario analysis, and this was adopted in the ERG base-case.

The company's corrected base-case ICERs (probabilistic) of pembrolizumab (with CAA) compared with SoC were £43,653 and £50,894 per QALY gained for cohort 1 and cohort 2 respectively. The cost effectiveness results were not robust to scenario and one-way sensitivity analyses conducted by the company. Scenario analyses indicated that response rates at week 12, the proportions of patients

receiving alloSCT, and the use of the MAIC instead of the naïve indirect comparison were major drivers of model results, mostly resulting in less favourable cost effectiveness estimates for pembrolizumab versus SoC.

The ERG incorporated various adjustments to the company's base-case. The ERG base-case resulted in ICERs (probabilistic) of pembrolizumab (with CAA) versus SoC of £64,186 and £78,696 per QALY gained for cohorts 1 and 2 respectively. For comparison, the deterministic ERG base-case ICERs were £61,705 and £73,594 per QALY gained for cohorts 1 and 2 respectively. The three most influential adjustments made by the ERG in its base-case for both cohorts were (in descending order) (1) the use of alternative utility values, (2) the use of the MSD survey only to estimate uptake of alloSCT instead of combining the MSD and BMS surveys, and (3) allowing alloSCT also in patients in the progressed disease state, in line with the MSD survey.

The ERG identified major and minor issues and uncertainties that affected the cost effectiveness analysis. Major issues and uncertainties are listed in the following. One major limitation was the company's model structure, which induced the implausible assumption that patients could only be eligible and receive alloSCT at 12 weeks after treatment start. The ERG deemed this implausible because response may, in reality, be obtained later than at 12 weeks and because, in practice, there is a lag between the decision to pursue alloSCT and the time at which the procedure is performed. The assumption lacked appropriate justification and deviated from how alloSCT was incorporated in TA462. The model is therefore a poor reflection of reality. Also, this model structure necessitated the differential fitting of parametric models to survival data for the pre- and post-12 week periods, inducing additional uncertainty. Furthermore, the company's model assumed that no patients would progress after receiving alloSCT. The impact of the limitations related to the model structure on model outcomes is unknown.

It should be noted that the appropriate approach for incorporating alloSCT in the model would have been to use time to alloSCT data directly from the main source of evidence. However, KEYNOTE-087 was not designed as a bridging study and poorly reflected clinical practice in the UK in terms of alloSCT uptake. The company therefore opted to inform alloSCT uptake conditional on response status at 12 weeks after treatment start through a UK clinician survey and then combined these survey results with the previously performed BMS survey results (from TA462). The ERG did not deem the combination of both surveys appropriate and considers there to be major uncertainty about the alloSCT uptake estimates. Furthermore, the elicited alloSCT uptake (from the MSD survey) for patients with progressed disease was ignored. Both, the combining of both surveys and ignoring alloSCT uptake in progressed disease patients, were shown in scenario analysis to be major drivers of cost effectiveness.

A major limitation was the use of single-arm evidence to inform treatment effectiveness. There was uncertainty whether the MAIC or the naïve indirect comparison should be used. The company provided both and the ERG, like the company, used the naïve indirect comparison in the base-case and the MAIC in scenario analysis. Furthermore, the ERG viewed the immaturity of the OS data from KEYNOTE-087 as a major limitation as this necessitated the use of post-alloSCT OS and utility estimates from alternative data sources, one of which was based on 13 patients only. [redacted], and the ERG considers that these may be informative for the present analysis. Furthermore, the company's method used for extrapolating OS post-alloSCT was deemed by the ERG to over-estimate OS, which significantly favoured pembrolizumab.

It is of note that the population used for the comparator was a mixed population of cohorts 1 and 2, that is, that did and did not receive autoSCT, derived from Cheah et al.[7] The Cheah et al population is more comparable with KEYNOTE-087 cohort 1 than with cohort 2 in terms of patient characteristics. The

use of this mixed comparator population likely resulted in comparisons of pembrolizumab with SoC that may be favourable and non-favourable for pembrolizumab in cohorts 1 and 2 respectively, but this could not be formally explored in scenario analysis.

Of further note, the economic model, and the evidence from KEYNOTE-087, rely on the assumption that treatment with pembrolizumab is capped at 24 months, which is inconsistent with its SmPC. It is unclear whether in UK clinical practice pembrolizumab would also be provided for a maximum of 24 months. The company and the ERG explored the impact of relaxing this assumption in scenario analysis.

Model extrapolations lack face and external validity. For example, the company claims that End of Life criteria can be considered fulfilled, however, their model predicts life year gains of 53 months on standard of care.

In exploratory analysis the ERG found that removing the 24-months cap on TTD had the largest impact on the ICERs in cohort 1 and a significant impact in cohort 2, and increased them to £78,992 and £79,284 per QALY gained for cohorts 1 and 2 respectively. The exploratory analysis with the largest impact in cohort 2 (ICER increased to £95,712) and the second largest impact in cohort 1 (ICER increased to £78,204) was the use of alternative assumptions when extrapolating post-alloSCT OS using data from Lafferty et al (2017). In cohort 2, the use of alternative parametric models for post-12 week PFS also substantially increased the ICERs to £87,401 and £90,152 per QALY gained when using the Gompertz and generalised gamma respectively, reflecting the significant uncertainty about extrapolating PFS in this model. The use of the MAIC instead of the naïve indirect comparison decreased the ICERs to £54,466 and £60,372 per QALY gained for cohorts 1 and 2 respectively. Assuming a lower post-alloSCT utility to explore the effect of the omission of a progressed disease health state post-alloSCT resulted only in small increases in the ICERs (by approximately £2,000 in both cohorts).

In conclusion, given that the ERG base-case ICERs are estimated to be substantially above £60,000 per QALY gained for both cohorts, with none of the scenarios resulting in ICERs below £50,000 per QALY gained, and the significant uncertainty induced by modelling choices and the use of single-arm studies with immature OS data, uncertainty around the cost effectiveness of pembrolizumab remains substantial.

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

In Section 5.3 the ERG's base-case was presented, which was based on various changes compared to the company's base-case. Tables 6.1 and 6.2 show how each individual change impacts the ICER in cohorts 1 and 2 respectively, plus the combined effect of all changes simultaneously. The analyses numbers in these tables correspond to the analyses numbers reported in Section 5.3. Furthermore, the exploratory analysis is presented in Tables 6.3 and 6.4 for cohorts 1 and 2 respectively (conditional on the ERG base-case). Appendix 1 contains technical details on the analyses performed by the ERG.

Table 6.1: ERG base-case cohort 1 (deterministic), pembrolizumab with CAA

	Technologies	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	Pembrolizumab ICER (£/QALY)
Company corrected base-case cohort 1	Pembrolizumab	£107,459	4.497			
	SoC	£52,017	3.223	£55,442	1.274	£43,511
Fixing errors (1)-(2)	Pembrolizumab	£107,459	4.496			
	SoC	£52,017	3.215	£55,442	1.282	£43,262
MSD survey only (3)*	Pembrolizumab	£105,128	4.072			
	SoC	£45,920	2.848	£59,208	1.224	£48,363
Time horizon 50 years (4)*	Pembrolizumab	£107,459	4.582			
	SoC	£52,017	3.275	£55,442	1.307	£42,412
Include monitoring costs post-alloSCT (5)*	Pembrolizumab	£110,298	4.496			
	SoC	£54,004	3.215	£56,294	1.282	£43,927
Alternative utility values (6)*	Pembrolizumab	£107,459	4.669			
	SoC	£52,017	3.617	£55,442	1.052	£52,705
Alternative pre-12 week OS distributions (7)*	Pembrolizumab	£107,496	4.499			
	SoC	£52,054	3.218	£55,442	1.282	£43,262
Proportion of alloSCT in PD state from MSD survey (8)*	Pembrolizumab	£107,934	4.524			
	SoC	£55,125	3.397	£52,809	1.127	£46,841
ERG base-case cohort 1 (combining adjustments 1-8)	Pembrolizumab	£107,998	4.460			
	SoC	£50,913	3.535	£57,085	0.925	£61,705

ERG = Evidence Review Group; ICER = incremental cost effectiveness ratio; QALY = quality-adjusted life year
* conditional on fixing errors (1) - (2)

Table 6.2: ERG base-case cohort 2 (deterministic), pembrolizumab with CAA

	Technologies	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	Pembrolizumab ICER (£/QALY)
Company corrected base-case cohort 2	Pembrolizumab	£93,732	4.072			
	SoC	£51,424	3.200	£42,308	0.871	£48,571
Fixing errors (1)-(2)	Pembrolizumab	£93,732	4.071			
	SoC	£51,424	3.193	£42,308	0.878	£48,178
MSD survey only (3)*	Pembrolizumab	£89,745	3.633			
	SoC	£45,464	2.835	£44,281	0.798	£55,478
Time horizon 50 years (4)*	Pembrolizumab	£93,732	4.149			
	SoC	£51,424	3.251	£42,308	0.897	£47,141
Include monitoring costs post-alloSCT (5)*	Pembrolizumab	£96,327	4.071			
	SoC	£53,378	3.193	£42,949	0.878	£48,908
Alternative utility values (6)*	Pembrolizumab	£93,732	4.309			
	SoC	£51,424	3.594	£42,308	0.714	£59,223
Alternative pre-12 week OS distributions (7)*	Pembrolizumab	£93,967	4.086			
	SoC	£51,607	3.208	£42,360	0.878	£48,236
Proportion of alloSCT in PD state from MSD survey (8)*	Pembrolizumab	£94,579	4.120			
	SoC	£54,466	3.371	£40,113	0.750	£53,508
ERG base-case cohort 2 (combining adjustments 1-8)	Pembrolizumab	£93,095	4.118			
	SoC	£50,609	3.541	£42,486	0.577	£73,594

ERG = Evidence Review Group; ICER = incremental cost effectiveness ratio; QALY = quality-adjusted life year
* conditional on fixing errors (1) - (2)

Table 6.3. Exploratory analysis conditional on ERG base-case cohort 1 (deterministic), pembrolizumab with CAA

	Technologies	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	Pembrolizumab ICER (£/QALY)
Company corrected base-case cohort 1	Pembrolizumab	£107,459	4.497			
	SoC	£52,017	3.223	£55,442	1.274	£43,511
ERG base-case cohort 1	Pembrolizumab	£107,998	4.460			
	SoC	£50,913	3.535	£57,085	0.925	£61,705
Alternative distributions (1.a)	Pembrolizumab	£107,552	4.361			
	SoC	£50,937	3.540	£56,615	0.821	£68,966
Use of MAIC (2)	Pembrolizumab	£107,998	4.460			
	SoC	£47,997	3.359	£60,001	1.102	£54,466

	Technologies	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	Pembrolizumab ICER (£/QALY)
No 24-months cap on TTD (3)	Pembrolizumab	£123,990	4.460			
	SoC	£50,913	3.535	£73,077	0.925	£78,992
Lower post-alloSCT utility (4)	Pembrolizumab	£107,998	4.346			
	SoC	£50,913	3.446	£57,085	0.900	£63,420
Alternative OS post-alloSCT assumption (5)	Pembrolizumab	£107,030	3.558			
	SoC	£50,157	2.830	£56,873	0.727	£78,204
ERG = Evidence Review Group; ICER = incremental cost effectiveness ratio; QALY = quality-adjusted life year						

Table 6.4. Exploratory analysis conditional on ERG base-case cohort 2 (deterministic), pembrolizumab with CAA

	Technologies	Total costs	Total QALYs	Incremental costs	Incremental QALYs	Pembrolizumab ICER (£/QALY)
Company corrected base-case cohort 2	Pembrolizumab	£93,732	4.072			
	SoC	£51,424	3.200	£42,308	0.871	£48,571
ERG base-case cohort 2	Pembrolizumab	£93,095	4.118			
	SoC	£50,609	3.541	£42,486	0.577	£73,594
Alternative distributions (1.a)	Pembrolizumab	£92,750	4.040			
	SoC	£50,698	3.558	£42,052	0.481	£87,401
Alternative distributions (1.b)	Pembrolizumab	£92,556	3.995			
	SoC	£50,550	3.529	£42,007	0.466	£90,152
Use of MAIC (2)	Pembrolizumab	£93,095	4.118			
	SoC	£45,924	3.337	£47,171	0.781	£60,372
No 24-months cap on TTD (3)	Pembrolizumab	£96,380	4.118			
	SoC	£50,609	3.541	£45,771	0.577	£79,284
Lower post-alloSCT utility (4)	Pembrolizumab	£93,095	4.013			
	SoC	£50,609	3.453	£42,486	0.560	£75,835
Alternative OS post-alloSCT assumption (5)	Pembrolizumab	£92,204	3.287			
	SoC	£49,863	2.844	£42,341	0.442	£95,712
ERG = Evidence Review Group; ICER = incremental cost effectiveness ratio; QALY = quality-adjusted life year						

7. END OF LIFE

According to the NICE criteria for End of Life, the following criteria should be satisfied:

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

According to the company there is not a valid estimate of OS for patients with RRcHL within UK clinical practice. However, based on their literature searches, the company estimate that OS ranges from 17.1 months to 19 months (CS, Table 51, page 129). In addition, in TA462 (Nivolumab for treating relapsed or refractory classical Hodgkin lymphoma) the committee "acknowledged that nivolumab did not unequivocally meet the criterion for short life expectancy but that it was plausible that the criterion could apply. It therefore agreed that on balance, nivolumab met the criterion for short life expectancy, and that it would take this into account in its decision-making."[6]

Regarding the second criterion, the company states that "As of March 21st 2017 [redacted] for Cohorts 1 and 2. However, the small number of deaths reported during the current follow-up period (15.9 months) indicates a substantially longer median survival than that offered by current therapies. The OS rate at 15 months in cohort 1 and 2 was reported using Kaplan-Meier estimates at [redacted], respectively.[56, 57]". Based on the company's economic model base case, the company predictions are 74 months for pembrolizumab and 53 months for SoC; therefore, the increment is 21 months in cohort 1 (ERG BC: Pembrolizumab LYs: 5.968 in months 71.616; SoC LYs: 4.761 in months 57.132). For cohort 2, the model predicts 67 months for pembrolizumab and 52 months for SoC; therefore, the increment is 15 months (ERG BC Pembrolizumab LYs: 5.517 in months 66.204; SoC LYs: 4.767 in months 57.204).

Overall, the ERG believes that the second criterion is more likely to be met. Regarding the first criterion, there is considerable uncertainty.

8. OVERALL CONCLUSIONS

8.1 Statement of principal findings

The company did not identify any randomised controlled trials of pembrolizumab and its comparators in patients with classical Hodgkin Lymphoma who have either received autoSCT and BV or BV alone due to autoSCT being unsuitable. One ongoing, single arm study of the efficacy and safety of pembrolizumab was identified (KEYNOTE-087) and this formed the basis of the submission. KEYNOTE-087 includes 150 patients (14 UK patients) relevant to this appraisal. It covers both cohorts of interest (Cohort 1: people with relapsed or refractory cHL who have received autologous stem cell transplant and brentuximab vedotin and Cohort 2: patients who have received brentuximab vedotin when autologous stem cell transplant is not a treatment option). The company presented data based on a median follow up of 15.9 months. The median time on treatment was [redacted] days for Cohort 1 and [redacted] days for Cohort 2.

The primary outcome of KEYNOTE-087 was overall response rate (ORR) as assessed by independent committee. ORR was 75.4% in Cohort 1 and 66.7% in Cohort 2. Median progression free survival (PFS) in Cohort 1 was 16.7 months (11.2 to NR). In cohort 2 it was 11.1 months (7.6 to 13.7). Median overall survival (OS) was [redacted]. At 12 months survival was [redacted] in Cohort 1 and [redacted] in Cohort 2. In cohort 1 [redacted] of patients had one or more adverse events. In Cohort 2 [redacted] of patients had one or more adverse events. The company noted that most AEs were low grade ([redacted] and [redacted] grades 3 to 5 in Cohort 1 and 2 respectively). In Cohort 1 [redacted] of AEs were classed as serious and in Cohort 2 [redacted]. The most common adverse events were pyrexia, cough, fatigue, diarrhoea and vomiting. The company conducted post-hoc analyses of response at 12 weeks to use in the comparison of clinical and cost effectiveness. The ERG noted that overall response rates were lower at 12 weeks than over the whole course of the trial ([redacted]

As KEYNOTE-087 did not have a comparator group the company identified a retrospective observational study from the literature (Cheah 2016 et al) to use as a comparator. This is a USA database study in which patients received the following types of therapy: investigational agent(s), gemcitabine, bendamustine, any other alkylator, BV retreatment, platinum based treatment, autoSCT or alloSCT, or other treatment. The company has not provided separate data for comparators; instead a combined data set has been provided for multiple comparators.

The company performed two types of analyses: a naïve indirect comparison between KEYNOTE-087 and Cheah and a matched adjusted indirect treatment comparison (MAIC) of the two studies.

Almost all results for PFS show a significant benefit for pembrolizumab versus SoC. Likewise, all results for ORR significantly favour pembrolizumab over SoC. Results of the naïve comparison are similar to MAIC. However, the results of the naïve comparison and MAIC are not reliable because they are likely to contain systematic error but it is not possible to estimate the size of the potential error. Both have major limitations and neither are fully reliable for decision making.

With regards to the health economic model submitted by the company, the ERG demonstrated that there was substantial uncertainty surrounding the ICERs and that alternative assumptions could change the ICER significantly. One major limitation was the company's model structure, which induced implausible assumption around the timing of alloSCT. The model was therefore considered a poor reflection of reality and likely to over-estimate cost effectiveness of pembrolizumab. There also remains substantial uncertainty about the uptake of alloSCT.

The use of single-arm evidence to inform treatment effectiveness was viewed as a major limitation and there was uncertainty about whether the MAIC or the naïve indirect comparison should be used. Furthermore, the ERG viewed the immaturity of the OS data from KEYNOTE-087 as a major limitation and the ERG considers that future data cuts may be informative for the present analysis. It is of note that the population used for the comparator was a mixed population of patients that did and did not receive autoSCT, which likely resulted in comparisons of pembrolizumab with SoC that may be favourable and non-favourable for pembrolizumab in cohorts 1 and 2 respectively. Of further note, the economic model, and the evidence from KEYNOTE-087, rely on the assumption that treatment with pembrolizumab is capped at 24 months, which is inconsistent with its SmPC. It is unclear whether in UK clinical practice pembrolizumab would also be provided for a maximum of 24 months. The substantial uncertainty in the evidence translates into model extrapolations that lack face and external validity. For example, the company claims that End of Life criteria can be considered fulfilled, however, their model predicts life year gains of 53 months on standard of care.

Apart from this, numerous issues were identified by the ERG. The ERG was able to adjust/correct some of these in its base-case. This resulted in probabilistic ICERs of pembrolizumab (with CAA) versus SoC of £64,186 and £78,696 per QALY gained for cohorts 1 and 2 respectively.

Additional sensitivity analyses were performed to examine the potential impact of alternative assumptions on the cost effectiveness estimates. The scenarios with the largest impact were alternative assumptions for extrapolating post-alloSCT OS (upward effect on the ICER), alternative survival models for extrapolating post-12 week PFS (upward effect on the ICERs), the use of the MAIC instead of the naïve comparison (downward effect on the ICERs) and removing the cap of 24 months on TTD (upward effect on the ICERs).

In conclusion, given that the ERG base-case ICERs are estimated to be substantially above £60,000 per QALY gained for both cohorts, with none of the scenarios resulting in ICERs below £50,000 per QALY gained, and the significant uncertainty induced by modelling choices and the use of single-arm studies with immature OS data, uncertainty around the cost effectiveness of pembrolizumab remains substantial.

8.2 Strengths and limitations of the assessment

The majority of searches for eligible studies in the CS were well documented and easily reproducible. Searches were carried out on all databases recommended in the NICE 2013 guide to the methods of technology appraisal sections 5.2.2 and 5.2.4. The clinical effectiveness strategies utilised recognised study design filters. Supplementary searches of conference proceedings and the NICE website and the WHO ICTRP trials database, were undertaken by the company in order to identify additional studies not retrieved by the main searches.

The clinical evidence is based on a well conducted, multicentre single-arm trial reflecting both cohorts of patients relevant to the decision problem. Outcomes assessed reflect the scope.

The main weakness is the lack of RCTs in this appraisal. Outcomes relating to pembrolizumab are based on a single arm trial. Comparisons with the comparators in the scope are problematic due to the availability of only one US study with a mix of different treatments. The naïve and matched adjusted comparisons conducted by the company have a number of limitations and represent a much weaker level of evidence than a RCT. Additionally progression-free survival and overall survival data are not fully mature.

Overall, the model is well built and transparent. The company reflected that pembrolizumab can be considered as a bridging treatment to alloSCT by incorporating alloSCT in the economic model. The company provided alternative data (for example derived from the MAIC) and alternative survival functions to enable exploratory analyses in the model.

AlloSCT was not appropriately reflected in the model, and there was substantial uncertainty about its uptake, as well as post-alloSCT survival and progression of patients. The use of single-arm evidence to inform treatment effectiveness was also viewed as a major limitation, inducing substantial uncertainty about relative treatment effectiveness. Furthermore, the ERG viewed the immaturity of the OS data from KEYNOTE-087 as a major limitation. [redacted] and may be informative for the present analysis. The assumptions made for extrapolating post-alloSCT overall survival significantly favoured pembrolizumab. The use of a mixed comparator population for both cohorts, that is, those patients that did and did not receive autoSCT, likely resulted in comparisons of pembrolizumab with SoC that may be favourable and non-favourable for pembrolizumab in cohorts 1 and 2 respectively. Another concern is the assumption that treatment with pembrolizumab is capped at 24 months, which favours pembrolizumab in this analysis.

8.3 Suggested research priorities

KEYNOTE-087 is an ongoing trial so more information will be available regarding uncertainties in progression-free and overall survival and other outcomes.

9. REFERENCES

[1] Merck Sharp & Dohme. *Pembrolizumab for treating relapsed or refractory classical Hodgkin's Lymphoma [ID: 1062]: Submission to National Institute of Health and Care Excellence. Single technology appraisal (STA)*, Sept 2017 [accessed 14.9.17]. 258p.

[2] Cancer Research UK. Hodgkin lymphoma: types [Internet]. 2014 [accessed 5.10.17]. Available from: http://www.cancerresearchuk.org/about-cancer/hodgkin-lymphoma/types

[3] Moskowitz AJ, Perales MA, Kewalramani T, Yahalom J, Castro-Malaspina H, Zhang Z, et al. Outcomes for patients who fail high dose chemoradiotherapy and autologous stem cell rescue for relapsed and primary refractory Hodgkin lymphoma. *Br J Haematol* 2009;146(2):158-63.

[4] National Institute for Health and Care Excellence. *Pembrolizumab for treating relapsed or refractory classical Hodgkin lymphoma: Final scope [Internet]*. London: NICE, March 2017 [accessed 13.9.17] Available from: https://www.nice.org.uk/guidance/gid-ta10130/documents/final-scope

[5] National Institute for Health and Care Excellence. *Brentuximab vedotin for treating CD30-positive Hodgkin lymphoma: NICE technology appraisal guidance 446 [Internet]*. London: NICE, 2017 [as cited in CS] Available from: https://www.nice.org.uk/guidance/ta446/resources/brentuximab-vedotin-for-treating-cd30positive-hodgkin-lymphoma-pdf-82604787983557

[6] National Institute for Health and Care Excellence. *Nivolumab for treating relapsed or refractory classical Hodgkin lymphoma: NICE technology appraisal guidance TA462 [Internet]*. London: NICE, 2017 [accessed 18.9.17] Available from: https://www.nice.org.uk/guidance/ta462/resources/nivolumab-for-treating-relapsed-or-refractory-classical-hodgkin-lymphoma-pdf-82604902197445

[7] Cheah CY, Chihara D, Horowitz S, Sevin A, Oki Y, Zhou S, et al. Patients with classical Hodgkin lymphoma experiencing disease progression after treatment with brentuximab vedotin have poor outcomes. *Ann Oncol* 2016;27(7):1317-23.

[8] Phillippo DM, Ades AE, Dias S, Palmer S, Abrams KR, Welton NJ. *NICE DSU technical support document 18: methods for population-adjusted indirect comparisons in submissions to NICE*. Sheffield: Decision Support Unit, ScHARR, University of Sheffield, 2016. 82p. Available from: http://scharr.dept.shef.ac.uk/nicedsu/wp-content/uploads/sites/7/2017/05/Population-adjustment-TSD-FINAL.pdf

[9] National Institute for Health and Care Excellence. *Pembrolizumab for treating relapsed or refractory classical Hodgkin's lymphoma [ID1062]. Clarification letter*. London: NICE, 2017

[10] Merck Sharp & Dohme. *Pembrolizumab for treating relapsed or refractory classical Hodgkin's lymphoma [ID1062]. Response to request for clarification from the ERG*: Merck Sharp & Dohme, 2017. 52p.

[11] Cooper K, Rose M, Harris P, Chorozoglou M, Picot J. *Nivolumab for treating relapsed or refractory classical Hodgkin lymphoma: a Single Technology Appraisal. ERG report TA462*: Southampton Health Technology Assessments Centre (SHTAC), 2017 [accessed 23.10.17] Available from: https://www.nice.org.uk/guidance/ta462/documents/committee-papers

[12] National Institute for Health and Care Excellence. *Pixantrone monotherapy for treating multiply relapsed or refractory aggressive non-Hodgkin's B‑cell lymphoma TA306 [Internet] [as cited in CS]*. London: NICE, 2014 Available from: https://www.nice.org.uk/guidance/TA306

[13] Food Drugs Administration. Pembrolizumab (KEYTRUDA) for classical Hodgkin lymphoma [Internet]. Silver Spring: FDA, 2017 [accessed 15.8.17]. Available from: https://www.fda.gov/Drugs/InformationOnDrugs/ApprovedDrugs/ucm546893.htm

[14] McGowan J, Sampson M, Lefebvre C. An evidence based checklist for the peer review of electronic search strategies (PRESS EBC). *Evid Based Libr Inf Pract* 2010;5(1):1-6.

[15] National Institute for Health and Care Excellence. *Single Technology Appraisal: specification for manufacturer/sponsor submission of evidence [Internet]*. London: NICE, 2012 [accessed 10.10.17]. 76p. Available from: http://www.nice.org.uk/Media/Default/About/what-we-do/NICE-guidance/NICE-technology-appraisals/Specification-for-manufacturer-sponsor-submission-of-evidence-June-2012.doc

[16] Scottish Intercollegiate Guidelines Network (SIGN). Search Filters [Internet]. Edinburgh: Healthcare Improvement Scotland, n.d. [accessed 2.11.17]. Available from: http://www.sign.ac.uk/search-filters.html

[17] Lefebvre C, Manheimer E, Glanville J. 6.4.9: Language, date and document format restrictions (Chapter 6: Searching for studies) [Internet]. In: Higgins JPT, Green S, editors. Version 5.1.0 [updated March 2011]: The Cochrane Collaboration, 2011 [accessed 11.10.16]. Available from: http://handbook.cochrane.org/

[18] Higgins JPT, Green S, eds. *Cochrane handbook for systematic reviews of interventions [Internet]*. Version 5.1.0 [updated March 2011]: The Cochrane Collaboration, 2011 [accessed 23.3.11]. Available from: http://handbook.cochrane.org/

[19] El Cheikh J, Massoud R, Haffar B, Fares E, Mahfouz R, Jisr T, et al. Bendamustine as a bridge to allogeneic transplant in relapsed/refractory Hodgkin lymphoma patients who failed salvage brentuximab vedotin postautologous peripheral blood stem cell transplantation. *Leuk Lymphoma* 2017;58(11):2745-7.

[20] Zinzani PL, Vitolo U, Viviani S, Corradini P, Motta G, Tani M, et al. Safety and efficacy of single-agent bendamustine after failure of brentuximab vedotin in patients with relapsed or refractory hodgkin's lymphoma: experience with 27 patients. *Clin Lymphoma Myeloma Leuk* 2015;15(7):404-8.

[21] Lafferty N, Anandram S, Lawes N, Usman M, Liebersbach S, Benn K, et al. Allogeneic stem cell transplantaion in patients with Hodgkin Lymphoma: a retrospective single centre case series (Abstract PO-075). Presented at 57th Annual Scientific Meeting of the British Society for Haematology; 27–29 March 2017; Brighton, UK. *Br J Haematol.* 2017;176(Suppl 1):72.

[22] Armand P, Shipp MA, Ribrag V, Michot JM, Zinzani PL, Kuruvilla J, et al. Programmed death-1 blockade with pembrolizumab in patients with classical Hodgkin lymphoma after brentuximab vedotin failure. *J Clin Oncol* 2016;34(31):3733-39.

[23] Alemayehu D. Assessing exchangeability in indirect and mixed treatment comparisons. *Comparative Effectiveness Research* 2011:51-55.

[24] Ishak KJ, Proskorovsky I, Benedict A. Simulation and matching-based approaches for indirect comparison of treatments. *PharmacoEcon* 2015;33(6):537-49.

[25] Cheson BD, Fisher RI, Barrington SF, Cavalli F, Schwartz LH, Zucca E, et al. Recommendations for initial evaluation, staging, and response assessment of Hodgkin and non-Hodgkin lymphoma: the Lugano classification. *J Clin Oncol* 2014;32(27):3059-68.

[26] Cheson BD, Pfistner B, Juweid ME, Gascoyne RD, Specht L, Horning SJ, et al. Revised response criteria for malignant lymphoma. *J Clin Oncol* 2007;25(5):579-86.

[27] Merck Sharp & Dohme. *Pembrolizumab for relapsed refractory classical Hodgkins Lymphoma [ID:1062]: Appendices to accompany company evidence submission. Single technology appraisal (STA)*, Sept 2017 [accessed 14.9.17]. 143p.

[28] National Institute for Health and Care Excellence. *Guide to the methods of technology appraisal 2013 [Internet]*. London: NICE, 2013 [accessed 25.10.17]. 93p. Available from: http://publications.nice.org.uk/pmg9

[29] Radford J, McKay P, Malladi R, Johnson R, Bloor A, Percival F, et al. Treatment pathways and resource use associated with recurrent Hodgkin lymphoma after autologous stem cell transplantation. *Bone Marrow Transplant* 2017;52(3):452-4.

[30] Doyle S, Lloyd A, Walker M. Health state utility scores in advanced non-small cell lung cancer. *Lung Cancer* 2008;62(3):374-80.

[31] Beusterien KM, Davies J, Leach M, Meiklejohn D, Grinspan JL, O'Toole A, et al. Population preference values for treatment outcomes in chronic lymphocytic leukaemia: a cross-sectional utility study. *Health Qual Life Outcomes* 2010;8:50.

[32] Lloyd A, Nafees B, Narewska J, Dewilde S, Watkins J. Health state utilities for metastatic breast cancer. *Br J Cancer* 2006;95(6):683-90.

[33] Nafees B, Stafford M, Gavriel S, Bhalla S, Watkins J. Health state utilities for non small cell lung cancer. *Health Qual Life Outcomes* 2008;6:84.

[34] Shingler SL, Swinburn P, Lloyd A, Diaz J, Isbell R, Manson S, et al. Elicitation of health state utilities in soft tissue sarcoma. *Qual Life Res* 2013;22(7):1697-706.

[35] Tolley K, Goad C, Yi Y, Maroudas P, Haiderali A, Thompson G. Utility elicitation study in the UK general public for late-stage chronic lymphocytic leukaemia. *Eur J Health Econ* 2013;14(5):749-59.

[36] National Institute for Health and Care Excellence. *Ticagrelor for preventing atherothrombotic events after myocardial infarction TA420 [Internet] [as cited in CS]*. London: NICE, 2016 Available from: https://www.nice.org.uk/guidance/ta420

[37] Department of Health. *Drugs and pharmaceutical electronic market information (eMit) [as cited in CS]*, 2017

[38] Woods B, Sideris E, Palmer S, Latimer N, Soares M. *NICE DSU Technical Support Document 19. Partitioned survival analysis for decision modelling in health care: a critical review [Internet]*, 2017 [accessed 31.8.17] Available from: http://www.nicedsu.org.uk

[39] MSD UK Ltd. *Data on file, Health Economist Expert Validation of cHL Cost Effectiveness Model Methodology [as cited in CS]*, 2017

[40] MSD UK M. *Clinicians Survey Results, MSD UK and MedeConnect*. MSD data on file, 2017

[41] Office for National Statistics (ONS). *National Life Tables: England 2013-15 2016 [Internet]*, 2016 [accessed 31.8.17] Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/lifeexpectancies/datasets/nationallifetablesenglandreferencetables

[42] Dolan P, Gudex C, Kind P, Williams A. *A social tariff for EuroQoL: results from a UK general population survey*. York: Centre for Health Economics, University of York, 1995

[43] Swinburn P, Shingler S, Acaster S, Lloyd A, Bonthapally V. Health utilities in relation to treatment response and adverse events in relapsed/refractory Hodgkin lymphoma and systemic anaplastic large cell lymphoma. *Leuk Lymphoma* 2015;56(6):1839-45.

[44] Ramsey SD, Nademanee A, Masszi T, Holowiecki J, Abidi M, Chen A, et al. Quality of life results from a phase 3 study of brentuximab vedotin consolidation following autologous haematopoietic stem cell transplant for persons with Hodgkin lymphoma. *Br J Haematol* 2016;175(5):860-867.

[45] Kurosawa S, Yamaguchi T, Mori T, Kanamori H, Onishi Y, Emi N, et al. Patient-reported quality of life after allogeneic hematopoietic cell transplantation or chemotherapy for acute leukemia. *Bone Marrow Transplant* 2015;50(9):1241-9.

[46] Joint Formulary Committee. British National Formulary [Internet]. London: BMJ Group and Pharmaceutical Press, June 2017. Available from: https://bnf.nice.org.uk/

[47] Department of Health. *NHS reference costs 2015 to 2016 [as cited in CS]*. London: Department of Health, 2016 Available from: https://www.gov.uk/government/publications/nhs-reference-costs-2015-to-2016

[48] National Institute for Health and Care Excellence. *Pembrolizumab for treating advanced melanoma after disease progression with ipilimumab TA357 [Internet] [as cited in CS]*. London: NICE, 2015 Available from: https://www.nice.org.uk/guidance/ta357

[49] National Institute for Health and Care Excellence. *Pembrolizumab for treating PD-L1-positive non-small-cell lung cancer after chemotherapy TA428 [Internet] [as cited in CS]*. London: NICE, 2017 Available from: https://www.nice.org.uk/guidance/ta428

[50] Brown T, Pilkington G, Bagust A, Boland A, Oyee J, Tudur-Smith C, et al. Clinical effectiveness and cost-effectiveness of first-line chemotherapy for adult patients with locally advanced or metastatic non-small cell lung cancer: a systematic review and economic evaluation. *Health Technol Assess* 2013;17(31):1-278.

[51] Department of Health. *NHS reference costs 2011 to 2012 [Internet] [as cited in CS]*. London: Department of Health, 2012 Available from: https://www.gov.uk/government/publications/nhs-reference-costs-financial-year-2011-to-2012

[52] Department of Health. *NHS reference costs 2013 to 2014 [Internet] [as cited in CS]*. London: Department of Health, 2014 Available from: https://www.gov.uk/government/publications/nhs-reference-costs-2013-to-2014

[53] Merck Sharp & Dohme. *Pembrolizumab for treating relapsed or refractory classical Hodgkin's Lymphoma [ID: 1062]: Submission to National Institute of Health and Care Excellence. Single technology appraisal (STA) [Version 2]*, Sept 2017 [received 16.11.17]. 258p.

[54] Perrot A, Monjanel H, Bouabdallah R, Quittet P, Sarkozy C, Bernard M, et al. Impact of post-brentuximab vedotin consolidation on relapsed/refractory CD30+ Hodgkin lymphomas: a large retrospective study on 240 patients enrolled in the French Named-Patient Program. *Haematologica* 2016;101(4):466-73.

[55] Kaltenthaler E, Carroll C, Hill-McManus D, Scope A, Holmes M, Rice S, et al. The use of exploratory analyses within the National Institute for Health and Care Excellence single technology appraisal process: an evaluation and qualitative analysis. *Health Technol Assess* 2016;20(26):1-48.

[56] Merck & Co. *Keytruda (MK-3475) relapsed or refractory classical hodgkin lymphoma HTA submission UK Report (July 2017). MSD data on file [as cited in CS]*, 2017

[57] Merck & Co. *KN-087 Efficacy Update Report (June 2017), A Phase II Clinical Trial of MK-3475 (Pembrolizumab in Subjects with Relapsed or Refractory (R/R) Classical Hodgkin Lymphoma (Database lock 21 March 2017). MSD Data on file*, 2017

Appendix 1: Technical details on the analyses performed by the ERG

ERG base-case			
1	Fixing error	AE disutility calculation	Outcome.calcs I6, X6
2	Fixing error	Remove patients' characteristics from the PSA	Parameters!D13:D15
3	Fixing violation	Use of MSD survey estimates only (rather than combined MSD and BMS surveys)	ClinicalData!S38:S40; ClinicalData!AP38:AP40
4	Fixing violation	Time horizon = 50 years	Control!C55
5	Fixing violation	Include TA462 monitoring costs post-alloSCT	Costs.calcs!R8,AK8
6	Matter of judgement	Alternative utility values	NonClinicalData C18:M22; Outcome.calcs J6, Y6
7	Matter of judgement	Alternative distributions for pre-12 weeks OS (exponential for cohort 1)	Survival!$I$32; ClinicalData!AJ13
7	Matter of judgement	Alternative distributions for pre-12 weeks OS (lognormal for cohort 2)	Survival!$I$32; ClinicalData!BG13
8	Matter of judgement	Proportion of alloSCT in PD health state from MSD survey	ClinicalData!S41; ClinicalData!AP41
ERG exploratory analyses			
1a)	Scenario	Cohort 1 & 2 : alternative distributions for for post-12 wks PFS (Gompertz)	Survival!$I$84; ClinicalData!V55, AC55, AS55, AZ55
1b)	Scenario	Cohort 2: alternative distributions for post-12 wks PFS b) generalised gamma	Survival!$I$84; ClinicalData!V55, AC55, AS55, AZ55
2	Scenario	Use of MAIC for HRs (PFS and OS) and response rates odds ratios	ClinicalData!U33:34; ClinicalData!AR33:34; ClinicalData!X63; ClinicalData!AE63; ClinicalData!AU63; ; ClinicalData!BB63
3	Scenario	Remove TTD cap at 24 months	NonClinicalData!G47
4	Scenario	Use lower post-alloSCT utility (i.e. the PD utility) to explore the impact of ignoring post-alloSCT PD	NonClinicalData!C20:C21
18	Scenario	Alternative assumption for post-alloSCT OS	ClinicalData!C77:D77 Survival!I117

Scenario analysis (5)

The ERG digitised the post-alloSCT OS KM estimates from CS Appendix 17 provided by the company and reconstructed IPD data. An alternative assumption was made regarding censoring, i.e. the ERG assumed censoring to occur after the last event for all but one remaining patients, instead of assuming all patients to be censored only at the end of follow-up. The ERG then fitted the Weibull and lognormal

curves to the generated IPD. In a validity test, the ERG found that it closely reproduced the company's results when assuming censoring at the end of follow-up only.