Pembrolizumab for Treating Relapsed or Refractory Classical Hodgkin Lymphoma: An Evidence Review Group Perspective of a NICE Single Technology Appraisal

Sabine E. Grimm¹ · Debra Fayter² · Bram L. T. Ramaekers¹ · Svenja Petersohn¹ · Rob Riemsma² · Nigel Armstrong² · Xavier Pouwels¹ · Willem Witlox¹ · Caro Noake² · Gillian Worthy² · Jos Kleijnen² · Manuela A. Joore¹

Published online: 21 March 2019
© The Author(s) 2019

Abstract

As part of its Single Technology Appraisal (STA) process, the National Institute for Health and Care Excellence (NICE) invited the manufacturer (Merck Sharp & Dohme; MSD) of pembrolizumab (Keytruda[®]) to submit evidence of its clinical and cost effectiveness for the treatment of patients with relapsed or refractory classical Hodgkin lymphoma (RRcHL) who did not respond to treatment with brentuximab vedotin. Kleijnen Systematic Reviews Ltd, in collaboration with Maastricht University Medical Centre+, was commissioned to act as the independent Evidence Review Group (ERG). The ERG produced a detailed review of the evidence for the clinical and cost effectiveness of the technology, based on the company's submission to NICE. According to the NICE scope, pembrolizumab was compared with single or combination chemotherapy. Comparisons were undertaken in two populations: patients who did and did not receive prior autologous stem cell transplant (autoSCT; populations 1 and 2, respectively). Despite it having been recommended by NICE in population 1 at the time the ERG received the company submission, nivolumab was not included as a comparator. No studies directly comparing pembrolizumab and its comparators were identified. One ongoing, single-arm study of the efficacy and safety of pembrolizumab (KEYNOTE-087) and one comparative observational study (Cheah et al., 2016) were used to inform the comparative effectiveness of pembrolizumab and standard of care (SoC), using indirect comparisons in both populations. Almost all analyses showed significant PFS and overall response rate benefits for pembrolizumab versus SoC, but due to being based on indirect comparison, were likely to contain systematic error. The economic evaluation therefore suffered from substantial uncertainty in any estimates of cost effectiveness. Furthermore, there was a lack of evidence on the uptake and timing of allogeneic stem cell transplant, and alternative assumptions had a significant impact on cost effectiveness. Immature survival data from KEYNOTE-087 exacerbated this issue and necessitated the use of alternative data sources for longer-term extrapolation of survival. Some issues identified in the company's analyses were amended by the ERG. The revised ERG deterministic base-case incremental cost-effectiveness ratios based on the company's second Appraisal Consultation Document response for pembrolizumab versus SoC (with a commercial access agreement) for populations 1 and 2 were £54,325 and £62,527 per quality-adjusted life-year gained, respectively. There was substantial uncertainty around these ICERs, especially in population 2. NICE did not recommend pembrolizumab as an option for treating RRcHL in population 1, but recommended pembrolizumab for use within the Cancer Drugs Fund in population 2.

✉ Sabine E. Grimm
Sabine.grimm@mumc.nl

¹ Department of Clinical Epidemiology and Medical Technology Assessment, School for Public Health and Primary Care (CAPHRI), Maastricht University Medical Centre+, P. Debyeilaan 25, PO Box 5800, 6202 AZ Maastricht, The Netherlands

² Kleijnen Systematic Reviews Ltd, 6 Escrick Business Park, Riccall Road, York YO19 6FD, UK

1 Introduction

The National Institute for Health and Care Excellence (NICE) is an independent organisation responsible for providing national guidance on promoting good health and preventing and treating ill health in priority areas with significant impact. Health technologies must be shown to be clinically effective and to represent a cost effective use of National Health Service (NHS) resources in order for NICE

Key Points for Decision Makers

This appraisal illustrates the increasing number of applications for marketing authorisation and reimbursement in cancer drugs based on single-arm studies and immature survival data. Guidelines on the circumstances in which the submission of non-randomised controlled trial evidence is acceptable would be useful. It would be helpful to establish at what point in the drug development process, and which level of data maturity reimbursement, applications become acceptable.

A recommendation subject to a commercial access agreement, and further data collection within the Cancer Drugs Fund, was made but the net value of this scheme was not formally assessed. Such formal assessment is important in ensuring the efficiency of the Cancer Drugs Fund.

Potential comparators not yet considered established practice or under appraisal at the time should be considered in the scope to avoid the omission of important comparators.

When a treatment enables subsequent treatment, or procedure, modelling assumptions around the timing of this procedure, subsequent health states and effectiveness evidence in the post-procedure period deserve in-depth scrutiny and may be sources of additional uncertainty. Alternative scenarios and flexible model structures should then be explored.

to recommend their use within the NHS in England. The NICE Single Technology Appraisal (STA) process usually covers new single health technologies within a single indication, soon after their UK market authorisation [1]. Within the STA process, the company provides NICE with a written submission, alongside a mathematical model that summarises the company's estimates of the clinical and cost effectiveness of the technology. This submission is reviewed by an external organisation independent of NICE (the Evidence Review Group [ERG]), which produces a report. After consideration of the company's submission (CS), the ERG report, and testimony from experts and other stakeholders, the NICE Appraisal Committee (AC) usually formulates preliminary guidance, the Appraisal Consultation Document (ACD), which indicates the initial decision of the AC regarding the recommendation (or not) of the technology. Stakeholders are then invited to comment on the submitted evidence and the ACD, after which a further ACD may be produced or a Final Appraisal Determination (FAD) issued, which is open to appeal.

The topic of this appraisal was pembrolizumab for the treatment of adult patients with relapsed or refractory classical Hodgkin lymphoma (RRcHL) who have not responded to autologous stem cell transplant (autoSCT) and brentuximab vedotin, or who are transplant-ineligible and have not responded to brentuximab vedotin. Full details of all relevant appraisal documents (including the appraisal scope, CS, ERG report [2], consultee submissions, ACD, FAD, and comments from consultees) can be found on the NICE website [3]. This paper summarises the ERG report [2], and highlights important methodological issues identified, which may be of benefit to future decision makers.

2 The Decision Problem

The underlying indication of this appraisal was classical Hodgkin lymphoma, which is a rare, localised or disseminated, malignant proliferation of cells of the lymphoreticular system, occurring mostly in lymph node tissues, spleen, liver, and bone marrow [4]. Classical Hodgkin lymphoma is the predominant subgroup of Hodgkin lymphoma and accounts for 95% of cases of the disease [5]. In 2014, there were 2106 new cases of Hodgkin lymphoma in the UK, with approximately half of diagnoses reported in patients aged 45 years and over, thus affecting many patients of working age [4].

For first-line therapy, chemotherapy alone or chemotherapy combined with radiotherapy is used in practice. However, between 15 and 30% of patients with Hodgkin lymphoma do not achieve remission with these treatments [6]. Patients who do not achieve remission may be offered chemotherapy and/or radiotherapy to enable autoSCT, which is potentially curative and effective in about 50% of patients [6]. 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. Brentuximab vedotin has recently been approved by NICE 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 [7, 8]. For those who do not respond to brentuximab vedotin, the prognosis remains poor with little options [4]. This is the point in the clinical pathway at which pembrolizumab is aimed. Regulatory approval by the European Medicines Agency (EMA) for the indication considered within this submission was granted on 2 May 2017. This stated that pembrolizumab as monotherapy is indicated for the treatment of adult patients with RRcHL who have not responded to autoSCT and brentuximab vedotin, or who are transplant-ineligible and have not responded to brentuximab vedotin. For those who are suitable, pembrolizumab may represent a bridge to allogeneic SCT (alloSCT), a potentially

curative treatment. Pembrolizumab is a highly selective humanised monoclonal antibody against the 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.

In terms of a comparator to pembrolizumab, there is no standard therapy after autoSCT and brentuximab vedotin [6]. Single or combination treatments including different chemotherapy regimens (some outside their marketing authorisation), such as gemcitabine, vinblastine and cisplatin, may be used, as listed in the NICE scope. The only other alternative is best supportive care (BSC).

3 Independent Evidence Review Group (ERG) Review

Kleijnen Systematic Reviews (KSR), in collaboration with the Maastricht University Medical Centre, acted as the ERG for this STA and reviewed the evidence on the effectiveness and cost effectiveness of pembrolizumab for this indication. In accordance with the process for STAs, the ERG and NICE had the opportunity to seek clarification on specific points in the CS [9], in response to which the company provided additional information [10]. The ERG also modified the company's decision analytic model to produce an ERG base-case and to assess the impact of alternative parameter values and assumptions on model results. Sections 3.1–3.6 below summarise this evidence, as well as the ERG's review of that evidence.

3.1 Clinical Effectiveness Evidence Submitted by the Company

The company did not identify any randomised controlled trials (RCTs) of pembrolizumab and its comparators in patients with classical Hodgkin lymphoma who have either received autoSCT and brentuximab vedotin or brentuximab vedotin 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 included 150 patients (14 UK patients) relevant to this appraisal. It covers both cohorts of interest (population 1: patients with RRcHL who have received autoSCT and brentuximab vedotin [$n=69$]; and population 2: patients who have received brentuximab vedotin when autoSCT is not a treatment option [$n=81$]). The company presented data based on a median follow-up of 15.9 months.

The primary outcome of KEYNOTE-087 was overall response rate (ORR) as assessed by an independent committee. The ORR was 75.4% (95% confidence interval [CI] 63.5–84.9) in population 1 and 66.7% (95% CI 55.3–76.8)

in population 2 over the course of the trial. Median progression-free survival (PFS) in population 1 was 16.7 months (95% CI 11.2 to not reached), and 11.1 months (95% CI 7.6–13.7) in population 2. Overall survival (OS) data from the trial were not mature.

As KEYNOTE-087 did not have a comparator group, the company identified a comparative observational study from the literature (Cheah et al.) [11]. This was a retrospective US database study in which patients received the following types of therapy: investigational agent(s), gemcitabine, bendamustine, any other alkylator, brentuximab vedotin retreatment, platinum-based treatment, autoSCT or alloSCT, or other treatment. The company did not provide separate data for comparators; instead a combined data set was provided for multiple comparators. The company performed two types of analyses: an unanchored naïve indirect comparison (IC) between KEYNOTE-087 and the study by Cheah et al., 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.

3.1.1 Critique of Clinical Effectiveness Evidence and Interpretation

The ERG noted that although KEYNOTE-087 was well conducted, it represents a low level of evidence. It is a phase II, single-arm, non-comparative study that, by its design, has serious limitations. The lack of a comparator arm means that the outcomes observed might not be a true reflection of the intervention as the role of natural history of the disease and the impact of patient characteristics are not taken into account. Furthermore, in an unblinded trial, knowledge of treatment received can lead to bias in the reporting of outcomes. These are just some of the limitations of a non-randomised trial. The ERG noted further limitations in applying the trial to clinical practice. These included the fact that survival data were immature. The trial had only 150 relevant participants and only a small number of patients were from the UK (14), therefore the trial might not reflect the UK population and setting.

The ERG also noted the limitations of the comparator single-arm study used in the submission as a proxy for standard care in the UK. However, in a previous appraisal of nivolumab (TA462) [12], the AC considered that “the Cheah study was the best available evidence for standard of care (SoC) and considered it appropriate for its decision-making, but overall the clinical effectiveness of nivolumab compared with SoC was highly uncertain because the comparator data may not fully represent UK clinical practice”.

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 study by Cheah et al. According to Decision Support Unit (DSU) Technical Support Document (TSD) report 18 [13], unanchored ICs (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 by Cheah et al. [11] and these are unlikely to be all relevant prognostic variables and effect modifiers. The ERG noted that the results were likely to contain systematic error, but it was not possible to estimate the size of the potential error. The ERG advised that the naïve IC and MAIC both had major limitations for decision making.

3.2 Cost-Effectiveness Evidence Submitted by the Company

Three systematic literature reviews (SLRs) were performed by the company with the aim of identifying all literature supporting the development and population of a model of

patients with RRcHL 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 versus pembrolizumab; (2) health-related quality of life (HRQoL) in the patient population; and (3) resource requirements and costs associated with treatment. No cost-effectiveness studies in patients with RRcHL were identified that met the inclusion criteria.

The company therefore built a de novo cohort state transition model with health states based on response, uptake of alloSCT, and survival (Fig. 1). This approach was adopted to reflect the expectation that pembrolizumab monotherapy would result in higher response rates than SoC, and hence be used as a 'bridge' to alloSCT. The model structure consisted 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, and a long-term component (after the first 12 weeks), which was modelled separately for patients who did and did not have alloSCT at 12 weeks. At 12 weeks, patients were allocated to alloSCT based on their response status and uptake probabilities

Fig. 1 Company's model structure for relapsed or refractory classical Hodgkin lymphoma. *alloSCT* allogeneic stem cell transplant, *CR* complete response, *PD* progressed disease, *PF* progression-free, *PR* partial response, *SD* stable disease

conditional on patients' response status. All patients in the model received alloSCT at this 12-week time point, without any lag. According to the company, these assumptions were made because alloSCT data from KEYNOTE-087 [14] were not reflective of UK clinical practice and no Kaplan–Meier (KM) data for time to alloSCT were available from the Cheah et al. study [11].

The model adopted the perspective of the NHS and Personal and Social Services (PSS) in England and Wales. The cycle length was 1 week to account for the length of treatment cycles. A half-cycle correction was applied. A time horizon of 40 years was adopted. All costs and quality-adjusted life-years (QALYs) were discounted at a rate of 3.5% per year.

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 not responded to autoSCT and brentuximab vedotin (population 1), and patients with RRcHL who are autoSCT ineligible and have not responded to brentuximab vedotin (population 2). Pembrolizumab monotherapy was 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 min every 3 weeks). The company assumed that pembrolizumab monotherapy would be provided for a maximum of 24 months.

The company only considered SoC as comparator in its base-case. SoC as considered by the company consisted of the following regimens: chemotherapy, bendamustine or investigational agents. The distribution of patients among these regimens was based on the distribution observed in the study by Cheah et al. [11]. Deviating from the scope, BSC was not included in the base-case, based on the belief that BSC use would be minimal as eligible patients are likely to receive therapy whenever feasible. BSC was added as a comparator in scenario analysis.

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. study [11]. An unanchored naïve IC was used to inform relative OS, PFS and response rates at week 12. The MAIC was only used in a scenario analysis. While KEYNOTE-087 data were used for separate analyses of the two populations, results were each compared with the whole Cheah et al. study cohort. Due to the company's model structure, treatment effectiveness and time to treatment discontinuation (TTD) were separately estimated for the pre and post 12-week period. 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 pooled 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. Post-alloSCT OS was based on the article by Lafferty et al. [15], by digitising KM curves, reconstructing individual patient-level data, and fitting survival models. Post-progression OS for patients who did not receive alloSCT was based on the study by Cheah et al. [11]. 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, but was capped at 24 months. TTD for SoC was assumed equivalent to PFS in the study by Cheah et al. Inclusion of adverse events (AEs) was in line with the previous Hodgkin lymphoma appraisal (TA 462) [12].

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, using response rates at 12 weeks to obtain progression-free health state utilities, and using response rates from the work of Lafferty et al. to calculate post-alloSCT utility. The company did not use the progressed disease (PD) utility score from KEYNOTE-087 and instead used a utility decrement from the paper by Swinburn et al. [16], which resulted in a lower PD utility score. To account for the possibility of acute graft-versus-host disease in the post-alloSCT health state, a disutility based on the study by Kurosawa et al. [17] was applied to the KEYNOTE-087 utility score in a proportion of patients (61.5% of patients).

The electronic market information tool (eMit [18]) was used to estimate drug acquisition costs of pembrolizumab and components of SoC. When these were unavailable, costs from the British National Formulary (BNF) [19] were used. Administration costs were obtained from NHS reference costs [20]. The list price of pembrolizumab 200 mg was £5260, but pembrolizumab was offered under a commercial access agreement (CAA). 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 non-alloSCT health states. For the post-alloSCT health state, a one-off cost based on the study by Radford et al. [21] was applied, which entailed alloSCT treatment costs, monitoring costs, costs of AEs, costs of subsequent treatment, and terminal care costs. No long-term costs were added. In the PD health state, BSC costs were applied as a one-off event and consisted of acquisition and administration costs of several treatments. Resource use and costs of AEs were based on NHS reference

costs and were applied in the model as one-off event costs. All costs were inflated to 2015–2016 values if necessary.

In the deterministic base-case analysis, total QALYs and life-years (LYs) gained, as well as total costs (with the CAA), were larger in the pembrolizumab treatment arm compared with UK SoC in both populations. Incremental costs mainly stemmed from differences in acquisition costs and alloSCT costs between pembrolizumab and SoC. After the company's corrections, pembrolizumab treatment resulted in deterministic (probabilistic, based on 1000 iterations) incremental cost-effectiveness ratios (ICERs) of £43,511 (£43,653) and £48,571 (£50,894) per QALY gained for populations 1 and 2, respectively. Sensitivity and scenario analyses resulted in significant changes to the ICERs in both populations.

3.2.1 Critique of Cost-Effectiveness Evidence and Interpretation

Searches were well reported and reproducible and unlikely to have missed any relevant studies. 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.

The model structure only allowed patients to have alloSCT at 12 weeks after starting treatment, which was viewed as a limitation as this meant ignoring responses that occurred at later time points. The alloSCT at 12 weeks assumption furthermore neglected any time lag caused by identifying a donor and scheduling the procedure. The ERG therefore considered alloSCT in the model to occur earlier than in clinical practice, resulting in post-alloSCT benefits being applied earlier, which would favour pembrolizumab. The exclusion of a post-alloSCT PD health state was not in line with evidence from Lafferty et al. [15], and also favoured pembrolizumab due to a larger number of patients undergoing alloSCT and therefore avoiding PD when treated with pembrolizumab.

The ERG considered the fact that the two populations were compared with the same comparator cohort with a mixed population from the study by Cheah et al., to result in comparisons of pembrolizumab with SoC that were favourable and non-favourable for pembrolizumab in populations 1 and 2, respectively. The exclusion of BSC as a comparator in the company's base-case was inconsistent with the scope. The ERG was most concerned that nivolumab, which was recently recommended by NICE [12] in population 1, and therefore a relevant comparator of pembrolizumab, was excluded from the model. The assumption that pembrolizumab monotherapy would be stopped after 24 months was inconsistent with the SmPC but in line with the KEYNOTE-087 protocol. It was unclear whether pembrolizumab

would also be provided for a maximum of 24 months in UK clinical practice. Removing this cap resulted in substantially increased ICERs for both populations, showing that the company's base-case might underestimate the cost incurred with the use of pembrolizumab if a 24-month stopping rule was not enforced in clinical practice. However, the ERG viewed the assumption as justified as it was in line with the clinical effectiveness evidence.

Treatment and relative treatment effectiveness estimates used in the model relied on the use of evidence from single-arm studies and an unanchored naïve IC, and were therefore highly uncertain. The use of the naïve IC instead of the MAIC was conservative. The combining of two 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 (which had been performed by different companies) to include the same clinical experts. Using only the company's own survey resulted in increased ICERs for pembrolizumab versus SoC. Post 12-week mortality data from KEYNOTE-087 were deemed immature by the company and the ERG agreed with this assessment. However, the ERG was also concerned about the use of the paper by Lafferty et al. for extrapolating post-alloSCT OS, given its small sample size ($n = 13$) and questionable generalisability to UK clinical practice, as well as limited information available in the KM plots, which did not provide patient numbers at risk at different time points. The company's method used for extrapolating OS post-alloSCT based on the KM plots in the paper by Lafferty et al. was deemed by the ERG to overestimate OS because the company appeared to have assumed no censoring in their analysis (Fig. 2). The ERG explored alternative assumptions around censoring in the tail of the KM estimates and produced its own extrapolated curves, which, based on visual inspection, appeared to make a better fit than the company's (Fig. 2). The company's approach was shown by the ERG to favour pembrolizumab by potentially overestimating post-alloSCT OS. 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 populations.

The mixed-effects model utilities using all observations instead of only the observations at week 12 that were provided in response to the clarification letter [10] were deemed by the ERG to make better use of the KEYNOTE-087 data. The ERG preferred estimating the PD utility from KEYNOTE-087, rather than from the work of Swinburn et al., as the estimates by Swinburn et al. did not adhere to the NICE reference case [1], being based on the time trade-off method and elicited from the general public. This utility estimate had previously been criticised in TA462 [12]. There was inconsistency in the calculation of the proportion of responders necessary for

Fig. 2 ERG approach versus the company's approach to estimating post-alloSCT overall survival. *alloSCT* allogeneic stem cell transplant, *ERG* Evidence Review Group

calculating utility values. Furthermore, the ERG preferred to use the utility score based on the study by Kurosawa et al. for the post-alloSCT health state, over the KEYNOTE-087 utility score, given that the KEYNOTE-087 data only included one patient who had undergone alloSCT and likely overestimated utility scores in the post-alloSCT population.

The ERG considered the assumption that all chemotherapy agents contributed equally to the mix of SoC in calculating costs to favour pembrolizumab. The likely underestimation of resource use and costs associated with alloSCT also favoured pembrolizumab.

Cost-effectiveness results were not presented for BSC in the base-case. The number of iterations (1000) in the probabilistic sensitivity analysis (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 [12] and the irreproducibility of model estimates used for external validation.

3.3 Addenda to the Original Company Submission and ERG Critique

In response to the first AC meeting, the company submitted additional evidence [22]. In this new submission, the company dismissed evidence by Eyre et al. [23] evaluating the efficacy of brentuximab vedotin in RRcHL patients in the transplant-naïve setting that could have informed analyses in population 2 instead of, or in addition to, data provided by Cheah et al. The company's grounds for dismissal of this evidence were differences in patient population between the study by Eyre et al. and KEYNOTE-087, small patient numbers in the relevant subpopulation ($n = 38$), and absence of KM estimates for the relevant subpopulation. However, the ERG noted that the first two arguments applied equally to the data from the study by Cheah et al. and that a potentially relevant source of evidence was therefore ignored from the analysis.

Furthermore, new cost-effectiveness models were submitted, in which some, but not all, of the ERG's preferred assumptions were incorporated. The most notable changes were the inclusion of a PD health state post-alloSCT (which was appropriately implemented, and increased ICERs) and the exploration of an alternative time point at which alloSCT was performed (24 weeks instead of 12 weeks after treatment initiation). The latter decreased the ICERs but the ERG considered it to be implemented inappropriately. One reason for this was that different survival models for PFS (the models with the worst statistical fit) were used compared with the original submission, without any justification. Most notably, the company changed the hazard ratio (HR) for OS in the pre-alloSCT period from an HR of 1 to an HR of 13.13. This change lacked external validity with the Cheah et al. data (model predictions: 24-week OS of 78% and 72% for populations 1 and 2, respectively, versus Cheah et al.: 26-week OS of 85%) and was neither explained nor could the analyses be reproduced because the necessary data were not provided.

In their response to the ACD after the second AC meeting, the company submitted additional evidence [22]. However, the company declined to provide the cost comparison with nivolumab in population 1, patients who had received prior autoSCT and brentuximab vedotin, which had been requested by the AC, stating that a cost comparison would not reflect any potential superiority in treatment effectiveness. However, the claim that pembrolizumab may be superior to nivolumab in terms of its treatment effectiveness was not explored. Updated model files included an updated price scheme, corrected distributions for estimating OS and PFS pre-alloSCT, and most of the ERG's preferences, presented for both the 12- and 24-week alloSCT assumption. However, the company continued to use the OS HR of 13.13 in their 24-week model, which the ERG therefore considered to be flawed. Alternative HRs were estimated by calibrating model outputs to the Cheah et al. data. This was not an evidence-based approach, but the ERG considered the resulting HRs to

be superior to the HR of 13.13 in terms of external validity and face validity, albeit highly uncertain. These alternative HRs increased the ICERs. An alternative PD utility value was estimated by using the midpoint between the data from Swinburn et al. and the KEYNOTE-087 utility score. This increased the ICERs compared with using the higher KEYNOTE-087 utility value because more patients transitioned to the PD post-alloSCT health state when treated with pembrolizumab than with SoC (while, in the original company's model, the absence of a post-alloSCT PD health state meant that lower PD utility scores would favour pembrolizumab). The ERG considered that utility estimate to be purely for illustrative purposes. In population 2, the company used evidence from the study by Eyre et al. [23] in a scenario, which substantially reduced ICERs in this population.

3.4 Additional Work Undertaken by the ERG

The ERG defined a new base-case that included multiple adjustments to the original base-case presented in the CS, based on identified errors, violations and alternative judgement [24]. This included correcting an error in the calculation of AE disutilities and excluding patient characteristics from the PSA. In terms of fixing violations, the ERG used only the company's own survey for informing probabilities of alloSCT uptake, extended the time horizon to a lifetime horizon (50 years instead of 40 years), and included long-term monitoring costs post-alloSCT. As matters of judgement, the ERG undertook multiple changes in the calculation of utilities, including using a mixed-effects model on all available observations, using the PD utility from KEYNOTE-087 instead of from the article by Swinburn et al. [16], and using the utility scores from the study by Kurosawa et al. [17], for the post-alloSCT group. Furthermore, the ERG changed the survival models used for pre 12-week OS in both populations to avoid overestimation of mortality. The ERG base-case analysis resulted in ICERs of pembrolizumab versus SoC that were substantially larger than those produced by the company. Based on the company's second ACD response (as reported in the FAD [25]), the revised ERG deterministic base-case ICERs, including alloSCT at week 12 (scenario with alloSCT at week 24), of pembrolizumab versus SoC were £54,325 (£45,829) and £62,527 (£42,501) per QALY gained in populations 1 and 2, respectively. The ERG considered that the alloSCT at week 24 ICERs should be interpreted with caution.

Furthermore, the ERG performed multiple exploratory analyses based on the ERG base-case. These analyses included the use of alternative parametric time-to-event models, use of the MAIC instead of the naïve treatment comparison, removal of the 24-month cap on TTD, use of a lower post-alloSCT utility to explore the impact of ignoring the post-alloSCT PD health state, and the use of alternative

assumptions (including some censoring before the end of the KM plots) to extrapolate post-alloSCT OS from the article by Lafferty et al. Almost all of these changes increased the ICERs substantially in both populations, with the notable exception of using the MAIC instead of the naïve treatment comparison. The company's population 2 scenario using alternative comparator effectiveness data from the study by Eyre et al. reduced the ICERs of this population group.

3.5 End-of-Life Criteria

According to the NICE criteria for end of life, the following criteria should be satisfied: (1) the treatment is indicated for patients with a short life expectancy, normally < 24 months; and (2) there is sufficient evidence to indicate that the treatment offers an extension to life, normally of at least an additional 3 months, compared with current NHS treatment. Overall, the ERG believed that the second criterion was more likely to be met. Regarding the first criterion, there was considerable uncertainty. The committee considered that while pembrolizumab did not unequivocally meet the criterion for short life expectancy, it is plausible that the criterion could apply.

3.6 Conclusions of the ERG Report

The main weakness of this appraisal was the lack of relevant non-RCT data. Outcomes relating to pembrolizumab were based on a single-arm trial. Comparisons with the comparators in the scope were 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 had a number of limitations and represent a much weaker level of evidence than an RCT. Additionally, PFS and OS data were not fully mature. KEYNOTE-087 is an ongoing trial therefore more information will be available in future regarding uncertainties in PFS and OS.

The revised ERG base-case ICERs were estimated to be above £50,000 per QALY gained, and the lower alternative ICERs based on the model implementation of alloSCT at 24 weeks were considered to be at least highly uncertain and potentially flawed. There remained many uncertainties, including the relative treatment effectiveness given the use of single-arm studies with immature OS data, the time point of alloSCT, uptake rates of alloSCT, longer-term and post-alloSCT survival, and the utility value in the PD health state. Some of these uncertainties were explored in scenario analyses and were mostly found to increase ICERs further. However, in population 2, there was additional uncertainty due to the indirectness of the comparator data, which resulted in the base-case ICER potentially being an overestimate. In the absence of a cost-effectiveness analysis comparing pembrolizumab with nivolumab, there was further uncertainty

about whether pembrolizumab was as (cost-)effective as nivolumab. These ICERs, their associated uncertainty and the uncertainty about whether end-of-life criteria were fulfilled resulted in substantial uncertainty about pembrolizumab being cost effective for patients with RRcHL in both populations—those who had not responded to, and those ineligible for, autoSCT after brentuximab vedotin treatment.

4 Key Methodological Issues

The clinical effectiveness of pembrolizumab submitted in this appraisal was based on one phase II, single-arm, non-comparative study that, by its design, had serious limitations. The outcomes observed might not be a true reflection of the intervention as the role of natural history of the disease and the impact of patient characteristics were not taken into account. Furthermore, in an unblinded trial, knowledge of treatment received can lead to bias in the reporting of outcomes. Methodological guidance is needed to deal with non-RCT evidence [26]. Guidelines on the circumstances in which the submission of non-RCT evidence is acceptable would also be helpful.

The main comparative study was a US observational study with a range of different treatments both within and outside the NICE scope. The ERG identified problems with compatibility of the two studies in the CS regarding baseline characteristics and the methods of outcomes assessment. In the MAIC, the company adjusted for potential confounding variables so that the KEYNOTE-087 study more closely resembled the study by Cheah et al. According to DSU report 18 [13], unanchored ICs (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 by Cheah et al. [11], 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.

When a treatment enables subsequent treatments or procedures, as is the case with pembrolizumab that potentially enables patients to receive allogeneic stem cell transplants, modelling assumptions around the timing of this procedure, subsequent health states, and effectiveness evidence in the post-procedure period deserve in-depth scrutiny and may be sources of additional uncertainty. In particular, the choice of one time point may not be suitable and alternative time points should be explored in scenarios, or more flexible model structures be considered.

When health-related quality-of-life data are collected in a trial, it is preferable to use all the available observations

from the trial, compared with using just the observations from one time point. This can be realised with a mixed-effects model, which should be reported with the model selection procedure, the covariates considered and used, and goodness-of-fit statistics.

In the absence of survival data, published KM plots can be digitised to re-create patient-level data and perform survival analysis. When the available figures do not show the patient numbers at risk at different time points, it is necessary to make assumptions about them and about any censoring, which is likely only possible in studies with very small sample sizes. Such analyses are very uncertain and do require an increased level of transparency in the reporting of methods, and should be accompanied by scenario analyses using alternative assumptions.

Finally, the consideration of pembrolizumab for reimbursement conditional on a CAA and further data collection within the Cancer Drugs Fund (CDF) should be informed by formal assessment of the value of such schemes. This could consist of establishing the expected cost effectiveness based on current evidence and the proposed price, the value forgone by making a recommendation without further research (expected value of perfect information [EVPI] or payer uncertainty burden [PUB]), and the value of any proposed research, which should be defined as unambiguously as possible in terms of outcomes, sample size, follow-up, and the timeframe at which it reports, compared with the cost of this research (reimbursement and data collection) [27–31].

5 National Institute for Health and Care Excellence Guidance

On 3 September 2018, NICE issued guidance that did not recommend pembrolizumab, within its marketing authorisation, as an option for treating RRcHL in adults who have had autoSCT and brentuximab vedotin, but did recommend pembrolizumab for use within the CDF as an option for treating RRcHL in adults who have had brentuximab vedotin and cannot have autoSCT, only if pembrolizumab is stopped after 2 years of treatment or earlier if the person has a stem cell transplant, or the disease progresses and the conditions in the managed access agreement for pembrolizumab are followed [25].

5.1 Consideration of Clinical- and Cost-Effectiveness Issues

This section summarises the key issues considered by the AC. The full list can be found in the FAD [25].

5.1.1 Comparators

Regarding comparator data, the AC stated that no data providing direct evidence for the clinical effectiveness of pembrolizumab compared with current standard care were available. The committee concluded that the study by Cheah et al. was the best available evidence for standard care at the time of the company's submission, particularly for population 1, but may not fully represent UK clinical practice. The committee welcomed the exploratory analyses based on the study by Eyre et al. [23] that the company provided for the third committee meeting, but noted that the company and ERG had concerns about using this study as a source of evidence for standard care.

5.1.2 Considerations of Clinical Effectiveness

The committee considered objective response rates and PFS assessed by blinded, independent central review from the most recent data-cut (March 2017) of the KEYNOTE-087 study. It noted that OS data from the trial were not mature.

The committee considered the unanchored naive IC and MAICs used in the company submission to compare pembrolizumab with SOC. The committee noted that these comparisons showed a beneficial effect for pembrolizumab for both of the outcomes included in the company's analysis (PFS and objective response rate). It also noted that these beneficial effects were generally higher in the MAIC than in the unanchored naive IC. The committee concluded that the ICs suggest that pembrolizumab has a beneficial effect on PFS and objective response rate, but that there was considerable uncertainty over the size of the effect and long-term outcomes.

5.1.3 Considerations of Cost Effectiveness

The AC considered that there was considerable uncertainty about whether the rates of uptake of alloSCT used in the models were an accurate reflection of transplant rates in UK clinical practice. The AC concluded that combining the results of the two surveys did increase the number of responses, although the combined number of responses was still small. However, there remained uncertainty about the validity and reliability of clinical predictions, as well as the potential duplication of clinicians in the combined survey. The AC further concluded that it was appropriate to assume that patients with PD do not have allogeneic stem cell transplants. However, the company's approach of omitting a PD state after allogeneic transplant was not considered by the AC to be appropriate or clinically plausible. The AC considered that stopping treatment with pembrolizumab at a maximum of 24 months in the models was appropriate. The AC concluded that there was considerable uncertainty about

the utility decrease that occurred when disease progressed, and that the actual value was likely to be between the company's and the ERG's base-case values.

The AC considered that the timing of alloSCT at 12 weeks into the model time horizon may be an underestimate and could potentially favour pembrolizumab because more patients treated with pembrolizumab would have alloSCT compared with SoC, and earlier transplant allowed them to benefit from an earlier point in time. It acknowledged that the alternative model assuming all transplants happen at 24 weeks allowed for exploring this uncertainty. However, the AC considered that the difference in OS between pembrolizumab and SoC was subject to uncertainty but was likely overestimated in this model version. Furthermore, the AC considered model choices for OS and PFS in the pre-alloSCT period in this model to introduce considerable uncertainty and considered the use of observed survival data in the model to have been preferable. The committee noted that how allogeneic stem cell transplant was incorporated in the models was a major driver of incremental QALYs for pembrolizumab compared with SoC, and concluded that the most plausible ICER was likely to fall between the values predicted by models using a fixed transplant time of 12 and 24 weeks.

The AC concluded that there was a lack of face validity between the modelled survival estimates for SoC (more than 2 years) and the clinical evidence, and the company's assertion that end-of-life criteria were met, which further added to the uncertainty about the results produced by the models. The committee concluded that, on balance, pembrolizumab met the end-of-life criteria.

The AC noted that the range of ICERs produced by the company's and ERG's 24-week and 12-week models were between £42,100 and £54,300 per QALY gained, respectively, but that these results were highly uncertain because of the total LYs predicted by the model, which were considered high, the uncertainties associated with the 24-week model, uptake rate, and timing of allogeneic stem cell transplant. The committee concluded that because of the substantial uncertainty associated with the model results, the ICERs for population 1 remained highly uncertain. The AC concluded that, in the absence of a cost comparison with nivolumab, it could only base its estimate of cost effectiveness for pembrolizumab in population 1 on the analyses comparing it with SoC, and that the results of these analyses were highly uncertain. The committee concluded that because of the substantial uncertainty associated with the model results, including the total LYs generated by the model, it was unable to predict the most plausible ICER for population 2, but the extreme values from the company's and ERG's 24-week and 12-week models (i.e. between £37,000 and £62,500 per QALY gained) reflected a plausible range in which the true ICER might fall. The committee concluded

that the estimates of cost effectiveness were too uncertain to recommend pembrolizumab for routine use.

The company considered pembrolizumab to be an innovative treatment. The AC noted the considerable unmet need for treatment in population 2 and that there was considerable uncertainty about the most plausible ICER for this population. It noted that time to alloSCT was a key driver of the cost-effectiveness estimates and there was considerable uncertainty about the true value. There was also uncertainty about whether the rates of alloSCT used in the models (which were based on clinician surveys) were an accurate reflection of transplant rates in UK clinical practice. The committee concluded that these were appropriate outcomes to collect data on, and this would reduce uncertainty in the cost-effectiveness estimate for population 2. It considered that OS for patients taking pembrolizumab would be a useful long-term outcome to measure. The committee concluded that pembrolizumab meets the criteria to be considered for inclusion in the CDF for population 2, and therefore recommended pembrolizumab for use within the CDF as an option for adults with RRcHL who have had brentuximab vedotin and cannot have autoSCT, if the conditions in the managed access agreement are followed.

6 Conclusions

This article describes the STA considering pembrolizumab for treating RRcHL in patients who have not responded to autoSCT and brentuximab vedotin, or who are transplant-ineligible and have not responded to brentuximab vedotin.

Pembrolizumab was not directly compared with nivolumab in this appraisal, although nivolumab had been recommended by NICE in one of the populations shortly prior to the start of this appraisal. This omission was made because nivolumab was still within the 90-day implementation period and was not yet considered established practice, and had in fact not been recommended at the time the scope was released. When requested by NICE later in the process, a cost comparison of pembrolizumab with nivolumab was not supplied by the company, resulting in additional uncertainty regarding the cost effectiveness of pembrolizumab compared with an increasingly used alternative treatment in this population. This highlights the importance of including all comparators in the scope and also including the possibility of additions of comparators under appraisal at the time.

This submission is another example of the increasing practice of submitting single-arm studies for marketing authorisation and reimbursement applications observed in cancer drugs over the past years [32, 33]. Methodological guidance is needed to deal with non-RCT evidence [26]. Guidelines on the circumstances in which the submission of non-RCT evidence is acceptable would also be helpful.

Furthermore, it would be helpful to establish at what point in the drug development process, and at which level of data maturity, reimbursement applications become acceptable.

In this case, the considerable uncertainty stemming from this sparse evidence base resulted in a recommendation subject to a CAA and further data collection within the CDF for one population. However, the value of such an evidence collection scheme was not formally assessed. A formal assessment should consist of establishing the expected cost effectiveness based on current evidence and the proposed price, the value forgone by making a recommendation without further research (EVPI or PUB), and the value of any proposed research, which should be defined as unambiguously as possible in terms of outcomes, sample size, follow-up, and the timeframe at which it reports, compared with the cost of this research (reimbursement and data collection) [27–31]. In this case, such formal assessment would have been challenging due to the committee's consideration of a range of ICERs across different model structures and the company's versus the ERG's preferences, but these challenges could have been overcome by methods for addressing structural uncertainty, such as model averaging [30, 34]. However, the EVPI and the value of specific research were not assessed, potentially resulting in suboptimal allocation of health care funding. As highlighted by Dickson et al. [33], the efficiency of the CDF could be ensured further if a formal step including the assessment of research and its costs was adopted.

Acknowledgements This summary of the ERG report was compiled after NICE issued the FAD. All authors have commented on the submitted manuscript and have given their approval for the final version to be published. The views and opinions expressed herein are those of the authors and do not necessarily reflect those of NICE or the Department of Health. Any errors are the responsibility of the authors.

Author Contributions Debra Fayter, Rob Riemsma, Nigel Armstrong and Jos Kleijnen critiqued the clinical-effectiveness data reported by the company. Caro Noake critiqued the literature searches undertaken by the company. Gillian Worthy critiqued the statistical analyses performed by the company. Sabine Grimm, Bram Ramaekers, Svenja Petersohn, Xavier Pouwels, Willem Witlox and Manuela Joore critiqued the mathematical model provided and the cost-effectiveness analyses submitted by the company. All authors were involved in drafting and commenting on the final document. Sabine Grimm acts as the guarantor of the manuscript. This summary has not been externally reviewed by PharmacoEconomics.

Compliance with Ethical Standards

Funding This project was funded by the National Institute for Health Research (NIHR) Health Technology Assessment Programme (Project Number 16/119/02). Please visit the HTA programme website for further project information (<https://www.nihr.ac.uk/funding-and-support/funding-for-research-studies/funding-programmes/health-technology-assessment/>).

Conflicts of interest Sabine E. Grimm, Debra Fayter, Bram L.T. Ramaekers, Svenja Petersohn, Rob Riemsma, Nigel Armstrong, Xavier Pouwels, Willem Witlox, Caro Noake, Gillian Worthly, Jos Kleijnen, and Manuela A. Joore have no conflicts of interest.

Open Access This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. National Institute for Health and Care Excellence. Guide to the methods of technology appraisal 2013. London: NICE; 2013. <http://publications.nice.org.uk/pmg9>. Accessed 25 Oct 2017.
2. Fayter D, Grimm S, Ramaekers B, Petersohn S, Riemsma R, Armstrong N, et al. Pembrolizumab for treating relapsed or refractory classical Hodgkin lymphoma: a Single Technology Assessment. York: Kleijnen Systematic Reviews Ltd; 2017. <https://www.nice.org.uk/guidance/ta540/documents/committee-papers>. Accessed 22 Nov 2018.
3. National Institute for Health and Care Excellence. Pembrolizumab for treating relapsed or refractory classical Hodgkin lymphoma: NICE technology appraisal guidance 540. London: NICE; 2018. <https://www.nice.org.uk/guidance/ta540>. Accessed 22 Nov 2018.
4. 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. <https://www.nice.org.uk/guidance/ta540/documents/committee-papers>. Accessed 14 Sep 2017.
5. Cancer Research UK. Hodgkin lymphoma: types. 2014. <http://www.cancerresearchuk.org/about-cancer/hodgkin-lymphoma/types>. Accessed 5 Oct 2017.
6. National Institute for Health and Care Excellence. Pembrolizumab for treating relapsed or refractory classical Hodgkin lymphoma: final scope. London: NICE; March 2017. <https://www.nice.org.uk/guidance/gid-ta10130/documents/final-scope>. Accessed 13 Sep 2017.
7. National Institute for Health and Care Excellence. Brentuximab vedotin for treating CD30-positive Hodgkin lymphoma: NICE technology appraisal guidance 446 [as cited in CS]. London: NICE; 2017. <https://www.nice.org.uk/guidance/ta446/resources/brentuximab-vedotin-for-treating-cd30positive-hodgkin-lymphoma-pdf-82604787983557>.
8. National Institute for Health and Care Excellence. Brentuximab vedotin for treating CD30-positive Hodgkin lymphoma: NICE technology appraisal guidance 524. London: NICE; 2018. <https://www.nice.org.uk/guidance/ta524/resources/brentuximab-vedotin-for-treating-cd30positive-hodgkin-lymphoma-pdf-82606840474309>. Accessed 11 Dec 2018.
9. National Institute for Health and Care Excellence. Pembrolizumab for treating relapsed or refractory classical Hodgkin's lymphoma [ID1062]. Clarification letter. London: NICE; 2017. <https://www.nice.org.uk/guidance/ta540/documents/committee-papers>. Accessed 22 Nov 2018.
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. <https://www.nice.org.uk/guidance/ta540/documents/committee-papers>. Accessed 22 Nov 2018.
11. 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.
12. National Institute for Health and Care Excellence. Nivolumab for treating relapsed or refractory classical Hodgkin lymphoma: NICE technology appraisal guidance TA462. London: NICE; 2017. <https://www.nice.org.uk/guidance/ta462/resources/nivolumab-for-treating-relapsed-or-refractory-classical-hodgkin-lymphoma-pdf-82604902197445>. Accessed 18 Sep 2017.
13. Phillippo D, Ades T, Dias S, Palmer S, Abrams K, Welton N. NICE DSU Technical Support Document 18: methods for population-adjusted indirect comparisons in submissions to NICE. Sheffield: NICE Decision Support Unit; 2016. <https://scharr.dept.shef.ac.uk/nicedsu/>. Accessed 24 Oct 2017.
14. Merck & Co. CSR: A Phase II Clinical Trial of MK-3475 (Pembrolizumab) in Subjects with Relapsed or Refractory (R/R) Classical Hodgkin Lymphoma (cHL) [database lock, June 2016]. MSD; 2016 (data on file; as cited in CS).
15. Lafferty N, Anandram S, Lawes N, Usman M, Liebersbach S, Benn K, et al. Allogeneic stem cell transplantation in patients with Hodgkin Lymphoma: a retrospective single centre case series [abstract no. PO-075]. Presented at the 57th Annual Scientific Meeting of the British Society for Haematology; 27–29 March 2017; Brighton. *Br J Haematol*. 2017;176 Suppl 1:72.
16. 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.
17. 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 Transpl*. 2015;50(9):1241–9.
18. Department of Health. Drugs and pharmaceutical electronic market information (eMit) [as cited in CS]. Department of Health; 2017.
19. Joint Formulary Committee. British National Formulary [as cited in CS]. London: BMJ Group and Pharmaceutical Press; June 2017. <https://bnf.nice.org.uk/>. Accessed 26 Oct 2017.
20. Department of Health. NHS reference costs 2015 to 2016 [as cited in CS]. London: Department of Health; 2016. <https://www.gov.uk/government/publications/nhs-reference-costs-2015-to-2016>.
21. 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 Transpl*. 2017;52(3):452–4.
22. Merck Sharp & Dohme. Pembrolizumab for treating relapsed or refractory classical Hodgkin's Lymphoma [ID: 1062]: additional analyses [Word document provided by the company]. 9 Feb 2018. Accessed 26 April 2018.
23. Eyre TA, Phillips EH, Linton KM, Kassam S, Gibb A, Allibone S, et al. Results of a multicentre UK-wide retrospective study evaluating the efficacy of brentuximab vedotin in relapsed, refractory classical Hodgkin lymphoma in the transplant naive setting. *Br J Haematol*. 2017;179(3):471–9.
24. 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.
25. National Institute for Health and Care Excellence. Pembrolizumab for treating relapsed or refractory classical Hodgkin lymphoma: final appraisal document. London: NICE; July 2018. <https://www.nice.org.uk/guidance/ta540/documents/committee-papers>. Accessed 22 Nov 2018.

- [nice.org.uk/guidance/ta540/documents/final-appraisal-determination-document](https://www.nice.org.uk/guidance/ta540/documents/final-appraisal-determination-document). Accessed 22 Nov 2018.
26. Anderson M, Naci H, Morrison D, Osipenko L, Mossialos E. A review of NICE appraisals of pharmaceuticals 2000–2016 found variation in establishing comparative clinical effectiveness. *J Clin Epidemiol*. 2019;105:50–9.
 27. Garrison LP Jr, Towse A, Briggs A, de Pouvourville G, Grueger J, Mohr PE, et al. Performance-based risk-sharing arrangements—good practices for design, implementation, and evaluation: report of the ISPOR good practices for performance-based risk-sharing arrangements task force. *Value Health*. 2013;16(5):703–19.
 28. Walker S, Sculpher M, Claxton K, Palmer S. Coverage with evidence development, only in research, risk sharing, or patient access scheme? A framework for coverage decisions. *Value Health*. 2012;15(3):570–9.
 29. Chalkidou K, Lord J, Fischer A, Littlejohns P. Evidence-based decision making: when should we wait for more information? *Health Aff (Millwood)*. 2008;27(6):1642–53.
 30. Grimm SE, Strong M, Brennan A, Wailoo AJ. The HTA risk analysis chart: Visualising the need for and potential value of managed entry agreements in health technology assessment. *Pharmacoecon*. 2017;35(12):1287–96.
 31. Eckermann S, Karnon J, Willan AR. The value of value of information: best informing research design and prioritization using current methods. *Pharmacoecon*. 2010;28(9):699–709.
 32. Davis C, Naci H, Gurpinar E, Poplavska E, Pinto A, Aggarwal A. Availability of evidence of benefits on overall survival and quality of life of cancer drugs approved by European Medicines Agency: retrospective cohort study of drug approvals 2009–13. *BMJ*. 2017;359:j4530.
 33. Dickson R, Boland A, Duarte R, Kotas E, Woolacott N, Hodgson R, et al. EMA and NICE appraisal processes for cancer drugs: current status and uncertainties. *Appl Health Econ Health Policy*. 2018;16(4):429–32.
 34. Bojke L, Claxton K, Sculpher M, Palmer S. Characterizing structural uncertainty in decision analytic models: a review and application of methods. *Value Health*. 2009;12(5):739–49.