in collaboration with:

Pembrolizumab for treating relapsed or refractory classical Hodgkin lymphoma - Addendum

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	06/02/2018

1

The company provided additional evidence following the first appraisal committee meeting. The ERG was asked to validate the additional work and comment on the impact of the amendments to the model.

The use of alternative data to inform cohort 2 analysis

The company have considered but eventually not taken into account the new evidence by Eyre et al (2017)[1] that could have informed the cohort 2 analysis (i.e. patients that did not have autologous SCT). The reasons for not taking this evidence into account included that patients in Eyre et al (2017)[1] are less heavily pre-treated than in KEYNOTE-087, that patients appear to be less far advanced in their disease course in Eyre et al compared with KEYNOTE-087, and that patient numbers relevant to the decision problem (based on their inability to receive autologous SCT; autoSCT) were considered to be small at n=30 by the company. In the absence of patient characteristics reported for this sub-population alone, and in the absence of further information on the subsequent intervention received and Kaplan-Meier estimates, the company chose not to include the data by Eyre et al (2017) in their cost effectiveness model.

The ERG remains unconvinced that the data reported in Eyre et al (2017)[1] could not be used to provide better estimates for cohort 2, i.e. those patients who have received brentuximab vedotin (BV) when autoSCT is not a treatment option. The ERG had highlighted before that the use of Cheah et al (2016)[2] for comparative evidence in cohort 2 was questionable due to the mixed population of patients receiving and not receiving autoSCT and the differences in baseline characteristics between the population in cohort 2 in KEYNOTE-087 and Cheah et al regarding age, ECOG score, Baseline B symptoms, Haemoglobin, Lymphocytes, Albumin, White cell count and Bulky Lymphadenopathy. The company's argument of a small relevant patient population in Eyre et al (2017) would also apply to Cheah et al (2016), where only n=27, that is fewer patients than in Eyre et al, did not undergo autoSCT. The sample size in Eyre et al (2017) could even be increased to n=38, if patients who received no further treatment were to be considered. Furthermore, since the company opted for a naïve comparison instead of a matched adjusted indirect comparison, it can be questioned whether the absence of patient characteristics in the sub-population hampers the usefulness of the data for the analysis. Whilst the absence of KM estimates for the relevant sub-population is a limitation, the ERG considers that the data collected by Eyre et al (2017) may present a relevant source of information that was not used in this analysis.

The company's newly submitted models

Upon the committee's request and ERG's recommendations, the company re-submitted two new economic models:

(1) the company's original corrected base-case model, but with the inclusion of a progressed disease health state after alloSCT (and two corrected technical errors identified by the ERG)

(2) the same model as above (model (1)), but with the implementation of an alternative time point at which patients would undergo allogeneic stem cell transplant (24 weeks instead of 12 weeks after treatment start)

The company disregarded the other changes made in the ERG base-case,[3] which included six further amendments to the model, some of which significantly increased the ICER and included the fixing of violations, such as the omission of long-term monitoring costs after alloSCT and the combination of two different surveys to inform the alloSCT uptake rates. The company explored some, but not all of these

2

amendments in their scenario analysis. Furthermore, the company made additional changes to model (2) by altering the distributions used for estimating PFS and the hazard ratio for OS, as well as amending response rates, odds ratios for response rates, utility values and estimates of time on treatment (see below for a more detailed description).

Model (1) – including a progressed disease health state after alloSCT

The ERG considers that the newly submitted model file (1) (when the changes made by the company are disabled) produce ICERs close enough to those produced by the ERG in Tables 6.1 and 6.2 of the original ERG report (when errors (1) and (2) are corrected) to instil confidence in that this model file is similar enough to the original to assess the impact of introducing a progressed disease (PD) health state for patients post-alloSCT. It is of note that, compared to the company's original corrected base-case, ICERs have increased with the inclusion of a progressed disease health state post-alloSCT. This is not caused by the company's adoption of the ERG's error correction, as correcting for these errors had decreased the ICERs in both cohorts.

However, the ERG firmly believes that its other changes (3) to (7) to the base-case should have also been used in the calculation of the new base-case ICERs and would have driven up the ICERs much more substantially. These were only explored in the company's scenarios, although not all of the changes made by the ERG were implemented correctly by the company, and these scenarios were not implemented in the models for the ERG to be able to validate them. As a result, the company's claim that the ICERs never exceeded the threshold of £50,000 per QALY gained is highly misleading: if the company's changes were implemented using the ERG base-case, i.e. considering these amendments simultaneously and correctly, the resulting base-case ICERs would very likely be significantly above £50,000 per QALY gained for the alloSCT at 12-week model file (model (1)) and only very slightly below it for the alloSCT at 24-week model file (model (2)), when the company's preferred PFS models and hazard ratio for 0-24 week OS are used. No rationale was provided by the company for the omission of these ERG amendments.

Since the company opted not to provide the changes in the model file in which the ERG implemented their amendments, there is no easy way to implement the ERG base-case within the company's new scenarios and demonstrate that the company's Scenario 11 is indeed not reflective of the ERG's amendments. With Table 1, the ERG wishes to illustrate why the company's new ICERs would likely exceed £50,000 per QALY gained in model (1), if the ERG base-case had been appropriately considered. The company's original base-case ICER in cohort 1 was increased by approximately £18,000 per QALY gained (£25,000 in cohort 2) with all the ERG base-case amendments. If the ERG amendment (8), i.e. patients with progressed disease being able to receive an alloSCT is disabled (according to clinical opinion heard at the first Appraisal Committee meeting), the increase in the ICER would still be approximately £14,000 per QALY gained for cohort 1 (£21,000 in cohort 2). There is no evidence for these ERG amendments being substantially less influential in the company's newly submitted model (1), where the ICER is £45,033 and £50,353 per QALY gained for cohorts 1 and 2 respectively, which means that it is likely that these ICERs would significantly exceed £50,000 per QALY gained if the ERG base-case (1) to (7) amendments were adopted.

Beyond the ERG base-case, in the original ERG report, the ERG had also performed exploratory analyses to represent the substantial uncertainty about survival prognosis after alloSCT, alternative OS and PFS extrapolations for patients without alloSCT and the use of a matched adjusted indirect comparison (MAIC). The former two increased the ICERs further and substantially (by up to £17,000

3

and £22,000 per QALY gained for cohorts 1 and 2), whilst the latter reduced the ICERs by approximately £5,000 and £13,000 per QALY gained for cohorts 1 and 2.

The ERG did not consider that the company's scenario 11 was equivalent to the ERG's combined preferences as stated by the company. Unfortunately, the company had not provided the model files with their scenarios implemented, and in the short time, the ERG was unable to produce its entire base-case in the two new submitted model files. Of greatest concern was the company's scenario analysis 1, for which the company stated that the mixed model utility values were used. This did not fully capture all the adjustments made in the original ERG report amendment (6). The ERG therefore performed an analysis using the company's base-case model (1) and re-implemented its amendments to utility values, which, apart from the use of the mixed model utilities also included: the use of KEYNOTE-087 to inform the progressed disease utility instead of Swinburn et al,[4] the calculation of the PFS utility for patients with and without alloSCT was calculated based on the respective patient proportions, the post-alloSCT utility was obtained from Kurosawa et al.[5] This increased the company's new ICERs to £52,876 and £59,452 per QALY gained instead of £51,319 and £57,308 per QALY gained as in the company's Scenario 1.

In summary, the ERG considers that the approximate increase of the ICERs of £2,000 per QALY gained caused by the introduction of the progressed disease health state appears plausible. It is however noteworthy that the company's Scenario 11 (including the post-alloSCT health state) still produces lower ICERs than the ERG's base-case amendments (1)-(7) without the inclusion of the post-alloSCT health state, as it did not fully reflect the ERG's preferences. Based on the ERG's exploration of Scenario 1 that indicated that if the ERG utility amendments were full considered the ICERs would increase, the ERG considers that the ICERs would more likely be higher than the company's scenario 11 ICERs and may be closer to the ERG original base-case amendments (1)-(7) ICERs, likely with an addition of £2,000 per QALY gained (Table 1).

Table 1. Company's original, ERG's original, and company's new 12 week (model 1) base-case with PD post-alloSCT

	Technologies	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	Pembrolizumab ICER (£/QALY)
Company original corrected base-case cohort 1	Pembrolizumab	£107,459	4.497			
	SoC	£52,017	3.223	£55,442	1.274	£43,511
ERG original base-case cohort 1	Pembrolizumab	£107,998	4.460			
	SoC	£50,913	3.535	£57,085	0.925	£61,705
Company resubmission model (1) – PD post-alloSCT cohort 1	Pembrolizumab	£107,459	4.328			
	SoC	£52,018	3.097	£55,441	1.231	£45,033
Company resubmission model (1) – Scenario 1 cohort 1	Pembrolizumab	£107,459	4.740			
	SoC	£52,018	3.660	£55,441	1.080	£51,319
ERG new scenario based on	Pembrolizumab	£107,460	4.655			
	SoC	£52,018	3.607	£55,441	1.049	£52,876

4

	Technologies	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	Pembrolizumab ICER (£/QALY)
model (1) scenario 1 but with amendments to utilities – cohort 1						
Company resubmission model (1) – Scenario 11 cohort 1	Pembrolizumab	£108,530	4.501			
	SoC	£48,305	3.428	£60,225	1.072	£56,160
ERG original base-case amendments (1)-(7), without post alloSCT PD state cohort 1	Pembrolizumab	£107,460	4.437			
	SoC	£47,558	3.392	£59,902	1.046	£57,275
Company original corrected base-case cohort 2	Pembrolizumab	£93,732	4.072			
	SoC	£51,424	3.200	£42,308	0.871	£48,571
ERG original base-case cohort 2	Pembrolizumab	£93,095	4.118			
	SoC	£50,609	3.541	£42,486	0.577	£73,594
Company resubmission model (1) – PD post-alloSCT cohort 2	Pembrolizumab	£93,733	3.917			
	SoC	£51,425	3.077	£42,307	0.840	£50,353
Company resubmission model (1) – Scenario 1 cohort 2	Pembrolizumab	£93,733	4.375			
	SoC	£51,425	3.637	£42,307	0.738	£57,308
ERG new scenario based on model (1) scenario 1 but with amendments to utilities – cohort 2	Pembrolizumab	£93,733	4.296			
	SoC	£51,426	3.584	£42,308	0.712	£59,452
Company resubmission model (1) – Scenario 11 cohort 2	Pembrolizumab	£93,025	4.132			
	SoC	£47,958	3.432	£45,066	0.700	£64,353
ERG original base-case amendments (1)-(7), without post	Pembrolizumab	£92,057	4.074			
	SoC	£47,224	3.396	£44,833	0.678	£66,133

	Technologies	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	Pembrolizumab ICER (£/QALY)
alloSCT PD state cohort 2						
ERG = Evidence Review Group; ICER = incremental cost effectiveness ratio; QALY = quality-adjusted life year						

Model (2) – alloSCT at 24 weeks instead of 12 weeks and including a progressed disease health state after alloSCT

The ERG considers the results of model (2), where the company implemented an alternative time point at which alloSCT is performed (24 weeks instead of 12 weeks) to suffer from substantial uncertainty. The company reported ICERs for cohort 1 of £39,880 per QALY gained, and of £39,714 per QALY gained for cohort 2. However, the ERG questions some of the changes undertaken by the company to implement the 24 week time point at which alloSCT is performed in the model. First, the company changed the PFS distributions for the time up to alloSCT (0-24 weeks period). However, the original distributions were fitted to the entire study data and a change of distributions should therefore be obsolete. The company now selects the curves with the worst statistical fit for PFS 0-24 weeks in cohort 1 (exponential instead of the previously chosen and best-fitting log-logistic), and for PFS 0-24 weeks in cohort 2 (again the exponential had the worst statistical fit and was selected over the previously chosen and best-fitting generalised gamma). The ERG implemented the previously chosen log-logistic curve for 0-24 week PFS in cohort 1 (which made the best statistical fit) and found that the ICER increased to £43,724 per QALY gained (£4,000 increase). The previously chosen best-fitting generalised gamma for cohort 2, however, decreased the ICER to £38,845 per QALY gained. This change is influential for cohort 1 and not in accordance with NICE DSU TSD 19 (Table 2).

Furthermore, the company chose to implement a hazard ratio (HR) of 13.13 (95% CI (3.07-56.04)) instead of 1 (as used in the original submission) for the estimation of relative treatment effectiveness in terms of overall survival in the 0-24 week period. This HR was pooled for cohorts 1 and 2. The HR could not be reproduced by the ERG, as the data for this were not provided. The ERG's concerns about this HR are that it could not be reproduced because the necessary data were not presented in Cheah et al.[2], and that a mixed KEYNOTE-087 population is used for its estimation. Furthermore, the model predictions for 24 weeks OS for patients treated with SoC are not in line with what is observed in Cheah et al (OS of 78% and 72% at 24 weeks in the model for cohort 1 and 2 versus approximately 85% alive at 26 weeks in Cheah et al.).

If this HR was set back to 1 in the newly submitted model (2), the resulting ICERs would increase to £45,048 and £48,524 per QALY gained for cohorts 1 and 2 respectively, that is, without any of the ERG's preferences implemented. The ERG acknowledges that this is an extreme scenario. However, this analysis illustrates that the resulting ICERs are remarkably close to the original company base-case ICERs (see Table 2, the 24-week time point for alloSCT has increased the ICER in cohort 1, and slightly decreased it for cohort 2). It therefore appears that the hazard ratio of 13.13 is the main reason for the model (2) ICERs being considerably lower than the original company's ICERs, with the caveat that other changes made to the model may also have had upward and downward effects on the ICERs. Another observation related to this is that the effect of the new alloSCT time point on costs and QALYs is by far not as substantial for pembrolizumab as it is for on Standard of Care (SoC) costs and QALYs, as can be seen in

6

Table 2. This is likely caused by a much shortened survival in these patients compared to the 12 week model (model (1)), which in turn may be a result of the HR of 13.13.

Furthermore, response data were changed to 24 week response data based on observed data from KEYNOTE-087 and a naive comparison was used to estimate odds ratios for response at this time point. Again, these odds ratios could not be verified, since the Cheah data for the 24-week time point were not available.

New distributions were fitted to the post 24 week PFS data. The company chose the exponential distributions for PFS post 24 weeks in cohorts 1 and 2. In cohort 1, this was the distribution with the best statistical fit, but in cohort 2, the exponential only ranks 4th and 5th according to AIC and BIC respectively. The generalised gamma would have been the distribution with the best statistical fit and the Gompertz was ranked second, but these were unfortunately not considered in the base-case or in scenario analysis in model (2).

Time to treatment discontinuation (TTD) post 24 weeks was also estimated using the data from KEYNOTE-087 up to a maximum of 24 months, and new distributions fitted. The exponential distribution was chosen for both cohorts, which exhibited the best statistical fit for cohort 1, and only ranked 4th and 2nd for cohort 2 according to AIC and BIC respectively. The Weibull would have made the best statistical fit to estimate post 24 week TTD in cohort 2 but was not explored by the company.

Lastly, utility values were updated to 24 week utility values. The ERG had preferred the use of all available utility data by estimating them using a mixed model, but the company did not apply this in their newly submitted base-case. The ERG therefore explored this, along with the other changes it had made to the utility values, in an exploratory analysis (

7

Table 2). This showed that the use of the ERG's preferences for utility values drove the ICERs up slightly for cohort 1 and down for cohort 2.

The ERG wishes to highlight that if the ICERs presented by the company could be accepted, the introduction of a new time point at which alloSCT is performed would result in decreases of the ICER of approximately £4,000 and £9,000 per QALY gained in cohorts 1 and 2 respectively. If using the ERG base-case as a starting point, and the same changes could be applied, this would still leave the ICERs at £53,000 and £57,000 per QALY gained for cohorts 1 and 2 respectively, based on the ERG (1)-(7) amendments. However, due to the significant changes to the structure and parameters of the model, it is not entirely clear whether the ERG preferences would have the same effect as they had on the original model.

The ERG considers that the ICER for this model with ERG preferences incorporated is likely above the one presented in the company's Scenario 11 for cohort 1, due to the effect of using the ERG's preferences for the utility values, because the chosen HR and 0-24 weeks OS curves appear to under-estimate OS for patients treated with SoC, and because of effects of choosing an alternative distribution for 0-24 week PFS. For cohort 2, the use of ERG preferences in the utility estimation would decrease the ICER, but the HR in combination with the choice of 0-24 weeks OS curve appears to substantially under-estimate the OS for patients treated with SoC. Therefore, it is difficult to know where the true ICER might lie, also in light of substantial uncertainties in this cohort.

Table 2. Company's original, ERG's original, and company's new 24 week (model 2) base-case with PD post-alloSCT

	Technologies	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	Pembrolizumab ICER (£/QALY)
Company original corrected base-case cohort 1	Pembrolizumab	£107,459	4.497			
	SoC	£52,017	3.223	£55,442	1.274	£43,511
ERG original base-case cohort 1	Pembrolizumab	£107,998	4.460			
	SoC	£50,913	3.535	£57,085	0.925	£61,705
Company resubmission model (2) – cohort 1	Pembrolizumab	£106,051	3.612			
	SoC	£34,320	1.813	£71,730	1.799	£39,880
Company resubmission model (2) – Scenario 1 cohort 1	Pembrolizumab	£106,051	4.503			
	SoC	£34,320	2.538	£71,730	1.965	£36,505
ERG new scenario based on model (2) scenario 1 but with amendments to utilities – cohort 1	Pembrolizumab	£106,051	4.454			
	SoC	£34,320	2.523	£71,731	1.930	£37,161
Company resubmission model (2) Scenario 11 – cohort 1	Pembrolizumab	£106,721	4.317			
	SoC	£33,536	2.532	£73,195	1.784	£41,021
ERG new scenario based on model (2) alternative distribution PFS 0-24 wks – cohort 1	Pembrolizumab	£111,085	3.726			
	SoC	£40,901	2.121	£70,184	1.605	£43,724
ERG new scenario based on model (2) but with HR=1 – cohort 1	Pembrolizumab	£106,051	3.612			
	SoC	£37,520	2.091	£68,531	1.521	£45,048
Company original corrected base-case cohort 2	Pembrolizumab	£93,732	4.072			
	SoC	£51,424	3.200	£42,308	0.871	£48,571
ERG original base-case cohort 2	Pembrolizumab	£93,095	4.118			
	SoC	£50,609	3.541	£42,486	0.577	£73,594
	Pembrolizumab	£89,726	3.154			

	Technologies	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	Pembrolizumab ICER (£/QALY)
Company resubmission model (2) –cohort 2	SoC	£33,217	1.731	£56,509	1.423	£39,714
ERG new scenario based on model (2) scenario 1 but with amendments to utilities – cohort 2	Pembrolizumab	£89,726	4.011			
	SoC	£33,217	2.395	£56,510	1.616	£34,979
Company resubmission model (2) Scenario 11 – cohort 2	Pembrolizumab	£89,408	3.898			
	SoC	£35,134	2.795	£54,274	1.103	£49,220
ERG new scenario based on model (2) alternative distribution PFS 0-24 wks – cohort 2	Pembrolizumab	£87,462	3.069			
	SoC	£29,828	1.585	£57,634	1.484	£38,845
ERG new scenario based on model (2) but with HR=1 – cohort 2	Pembrolizumab	£89,726	3.154			
	SoC	£37,128	2.070	£52,599	1.084	£48,524
ERG = Evidence Review Group; ICER = incremental cost effectiveness ratio; QALY = quality-adjusted life year						

Conclusion

In conclusion, whilst the company has addressed some important structural uncertainty in their new models, it was unable to resolve and address the substantial uncertainties present in their economic model and overall submission. First, the company did not make use of new evidence that could be of value to inform the analysis in patients that did not have autologous SCT (cohort 2), which currently is informed by a mixed population study that has been criticised by the ERG in its original ERG report. Second, the introduction of a post alloSCT progressed disease health state increased the ICERs. Third, there are substantial questions relating to the implementation of the alternative time point of 24 weeks at which patients may receive alloSCT. These questions relate mainly to the use of a hazard ratio for overall survival prior to 24 weeks from a mixed population for both cohorts that could not be verified by the ERG, and the choice of distributions for estimating PFS both before the 24 week point and after, as well as time on treatment after 24 weeks that do not exhibit the best statistical fit and lack other justification. The full effects of this on model outcomes could not be assessed by the ERG. Fourth, the company did not implement their changes using the ERG base-case. Fifth, substantial uncertainties highlighted by the ERG remain unexplored. This includes the method for extrapolating post-alloSCT

overall survival, where alternative assumptions increased the ICERs by £17,000 and £22,000 per QALY gained for cohorts 1 and 2 respectively.

The ERG therefore considers the ICERs presented in the company's Scenario 11 for both analyses to be under-estimates compared to the ERG's preferences. Even though the direction of potential bias introduced by the company's amendments in the cohort 2 week 24 model is less clear, there remain substantial upward uncertainties also for cohort 2.

REFERENCES:

[1] 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):471-479.

[2] 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.

[3] 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

[4] 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.

[5] 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.