**NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE**

**SINGLE TECHNOLOGY APPRAISAL**

**APPEAL HEARING**

**Advice on Pembrolizumab for treating locally advanced or metastatic urothelial carcinoma after platinum-containing chemotherapy**

# Decision of the panel

**Introduction**

1. An appeal panel was convened on 23 June 2020 to consider an appeal against NICE’s final appraisal document (FAD), to the NHS, on pembrolizumab for treating locally advanced or metastatic urothelial carcinoma after platinum-containing chemotherapy. The appeal was conducted via Zoom.
2. The Appeal Panel consisted of:

|  |  |
| --- | --- |
| * Prof Alan Silman
 | Chair |
| * Mr Tom Wright
 | NICE Non-executive director |
| * Dr Biba Stanton
 | NHS representative |
| * Dr Mark Chakravarty
 | Industry representative |
| * Mr John Morris
 | Lay representative |

1. None of the members of the appeal panel had any competing interest to declare.
2. The panel considered an appeal submitted by Merck Sharp & Dohme (MSD), the company.
3. MSD was represented by:

|  |  |
| --- | --- |
| * Benjamin Bates
 | Health economics manager, MSD |
| * Grant Castle
 | Partner, Covington & Burling LLP |
| * Dr Simon Crabb
 | Associate Professor of Medical Oncology, University of Southampton |
| * Kalpana D’Oca
 | Team Leader, HTA & OR, MSD |
| * Keiron Hughes
 | Head of Strategic Pricing, MSD |

1. In addition the following individuals involved in the appraisal were present and available to answer questions from the appeal panel:

|  |  |
| --- | --- |
| * Dr Lindsay Smith
* Nicola Hay
 | Committee Chair TAC D (for this appraisal)Technical advisor |
| * Helen Knight
 | Programme director |
| * Linda Landells
 | Associate director |
| * Prof Gary McVeigh
* Daniel Gallacher
 | Committee memberERG representative |

1. The appeal panel’s legal adviser Stephen Hocking of DACBeachcroft LLP was also present.
2. Two members of the NICE appeals panel (Mr Alan Thomas and Mr Tony Heddon) were present as observers but did not participate in any of the discussions of the appeal panel, or in the decision-making.
3. Under NICE’s appeal procedures members of the public are admitted to appeal hearings and several members of the public and NICE staff observed the appeal via Zoom.
4. There are two grounds under which an appeal can be lodged:

**Ground One: In making the assessment that preceded the recommendation, NICE has:**

1. **Failed to act fairly**
2. **Exceeded its powers.**

**Ground Two: The recommendation is unreasonable in the light of the evidence submitted to NICE.**

1. The then Vice Chair of NICE (Dr Rima Makarem) in preliminary correspondence had confirmed that MSD had potentially valid grounds of appeal as follows:
* Ground 1a – NICE has failed to act fairly
	+ Ground 1a.1 – Fundamental differences with the approach taken in TA525 evidence an inconsistent methodology, in breach of the Methods Guide and the principle of procedural fairness
	+ Ground 1a.2 – It is procedurally unfair to have introduced paragraph 3.19 (“The costs of pembrolizumab are likely underestimated in the model”) in the FAD at a very late stage, without explanation or any opportunity to respond
* Ground 1b – NICE has exceeded its powers
	+ Ground 1b.1 – NICE has breached its legal obligations under human rights and equalities laws
* Ground 2 – the recommendation is unreasonable in the light of the evidence submitted to NICE
	+ Ground 2.1 – The Appraisal Committee’s assessment of duration of treatment effect and its effect on cost-effectiveness is illogical and unreasonable\*
	+ Ground 2.2 – The Appraisal Committee’s analysis of evidence from the clinical expert in paragraph 3.15 of the FAD is internally inconsistent and its conclusions are unreasonable\*
		- \*Agreed as valid points with 2.2 to be taken as an aspect of 2.1
	+ Ground 2.3 – The Appraisal Committee’s decisions that: (i) a range of possible ICERs from £48,518 to £70,520 applies; and (ii) the “most plausible” ICER for pembrolizumab is likely to be above £50,000, are unreasonable in light of the evidence submitted
	+ Ground 2.4 – The statement that “the costs of pembrolizumab are likely underestimated in the model” lacks meaningful explanation and evidence and it is unreasonable to have taken it into account for the purposes of assessing cost-effectiveness
	+ Ground 2.5 – The conclusion that new data from KEYNOTE-045 “shows the 2-stage method may not be appropriate, and the unadjusted method should also be taken into account” results from a flawed and unreasonable interpretation of the evidence
1. The appraisal that is the subject of the current appeal provided advice to the NHS on pembrolizumab for treating locally advanced or metastatic urothelial carcinoma after platinum-containing chemotherapy.

1. Before the appeal panel inquired into the detailed complaints the following made a preliminary statement: Grant Castle on behalf of MSD, and Dr Lindsay Smith on behalf of the appraisal committee.

**Appeal Ground 1a: In making the assessment that preceded the recommendation, NICE has failed to act fairly.**

**Appeal Ground 1a.1:** Fundamental differences with the approach taken in TA525 evidence an inconsistent methodology, in breach of the Methods Guide and the principle of procedural fairness

1. Grant Castle, for MSD, said that the committee had deviated from an established, reasonable approach in the case of this appraisal. In particular, there were clear similarities between this appraisal and the appraisal of atezolizumab (TA525). Therefore, if a different approach were to be taken in this appraisal, it should be clearly justified. He said that the key example of the difference in approach is the different treatment effect duration assumptions used in the two appraisals (3 years from the end of treatment with a two year treatment stopping rule in TA525, 3-5 years from the start of treatment in this appraisal).
2. Mr Castle said that section 6.2.16 of the NICE methods guide suggests that appraisal committees should try to take a consistent approach unless there are differences in the evidence. In this case, pembrolizumab has been shown to have an effect on overall survival, where atezolizumab has not, and there are longer follow-up data available for pembrolizumab. Therefore, the differences in the evidence would be in favour of assuming a longer rather than shorter duration of treatment effect for pembrolizumab.
3. Mr Castle said that MSD had raised this concern at several points throughout the appraisal but had never received a satisfactory explanation.
4. Benjamin Bates, for MSD, explained that with a cap on treatment effect duration at 3 years from the start of treatment, the modelling assumes that for patients still alive on active treatment at this time point their hazard rate suddenly jumps to that of the standard of care arm. The company’s position is that the lower chance of dying with pembrolizumab continues past three years.
5. Mr Bates said that the approach taken by the ERG in this appraisal was significantly different from that used in other immune-oncology technology appraisals.
6. Dr Simon Crabb, for MSD, said that the real clinical benefit of pembrolizumab is not the improvement in median survival but that a small group of patients have very long survival, even after stopping treatment. He also emphasised that the clinical trials of atezolizumab and pembrolizumab were extremely similar.
7. Prof Gary McVeigh, for NICE, explained that in 2017-2018 atezolizumab and pembrolizumab were considered in parallel by the committee (in TA525 and TA519) but at that stage the data were extremely immature. At that time there was no evidence for a prolonged treatment effect, and it was impossible to know how many patients might have a prolonged effect from treatment. The committee had agreed that a lifetime effect was not plausible. He agreed that it is not plausible that the benefit of treatment is instantaneously lost at a specific time point for all patients. On the other hand, the data support a waning of the treatment effect over time (even before treatment is stopped). Using a treatment effect duration cap was therefore a compromise. The options presented to the committee in TA525 were either a 5 or 7 year treatment effect duration (3-5 years after a two year stopping rule).
8. Prof McVeigh said that the reason atezolizumab was approved for routine practice, whereas pembrolizumab went into the Cancer Drugs Fund, was due to the commercial offer around atezolizumab’s acquisition. With that offer, the range of plausible ICERs for atezolizumab was well within what is considered cost-effective for an End of Life treatment. Because of this, the duration of treatment effect was not a key determinant of the decision.
9. Later in the hearing (during discussion of Appeal Point 2.3) Helen Knight, for NICE, also made the point that one difference between the two appraisals lay that in TA525 the estimated ICERs were well within the range that is considered cost-effective and thus the uncertainties around treatment effect duration in this appraisal (TA519) were more critical to decision making than in TA525.
10. In response to a question from the panel, Prof McVeigh said that he remains convinced that the decision on atezolizumab in TA525 was correct given the evidence available at the time and the value proposition, incorporating the commercial offer, put to the committee.
11. The panel asked whether the committee had considered a more conservative estimate of treatment effect duration of 3 years (1 year after the 2-year stopping rule) in TA525. Prof McVeigh replied that this was not one of the options presented to the committee. They preferred the most conservative of the options presented and noted the huge degree of uncertainty. He also commented that a 3-year treatment effect duration would not have changed the decision because of the value proposition.
12. Prof McVeigh, for NICE, said that in this current appraisal the committee had considered treatment effect durations of 3 or 5 years, and did not have a preference for 3 years. He pointed out that a 5 year effect from the start of treatment (for pembrolizumab) is not exactly equivalent to a 3 year effect after stopping treatment with a stopping rule of 2 years (for atezolizumab) as more than half of patients were already off treatment after 6 months.
13. In response to a question from the panel, Prof McVeigh said that the committee recognised that a small minority of patients may have a prolonged benefit from pembrolizumab.
14. In response to a question from the panel, Prof McVeigh said that it is difficult to compare the decision in this appraisal to that in TA525 because we now have two more years of data. An expectation of consistency was more relevant to the two decisions in TA525 and TA519 because these were taken in parallel.
15. Grant Castle, for MSD, said that he had still not heard an explanation for why a different approach was taken in this appraisal compared with TA525.
16. Benjamin Bates, for MSD, emphasised that the company considered a 5-year treatment effect duration to be conservative and preferred a lifetime treatment effect. Dr Simon Crabb, for MSD, gave a clinical opinion that the plausible duration of treatment effect was five years at a minimum.
17. Dr Lindsay Smith, for NICE, argued that the committee were not inconsistent in their approach. In both instances, they were presented with a range of options by the company and ERG. In both instances, they considered these options. At the time of TA525 there was very little evidence, and they chose the most conservative option of those presented. For the current appraisal, there was longer follow-up data available. But having longer follow-up data does not necessarily mean this data supports a longer treatment effect. The committee considered the options given to it and reached a different decision.
18. Helen Knight, for NICE, emphasised that committees are mindful of the need for consistency, but that NICE procedures do not allow committees to consider models from other appraisals in reaching their decision. The starting point for committees is always the submissions from the company and ERG, which differ between appraisals.
19. The appeal panel concluded that there is a biologically plausible rationale for thinking that the effect of pembrolizumab may continue after treatment is stopped, but that there is a great deal of uncertainty about the duration of treatment effect.
20. The panel agreed that appraisal committees should ensure, as far as possible, that their judgements regarding the cost-effective use of NHS resources are consistently applied between appraisals. The panel also noted consistently with its past decisions that this expectation must not be set too high. Ms Knight is correct that the evidence from one appraisal cannot be considered in another (at least, not unless it has been expressly included in that other appraisal). It is also right that any expectation of consistency between very different drugs or in different disease areas may be set only at a very general level. And it is also right that one committee does not bind another, or even itself for future appraisals, and past approaches may be departed from with appropriate reasoning.
21. The panel agreed that the judgement on duration of treatment effect was materially different between this appraisal and TA525, and that duration of treatment effect was one of the key drivers of the decision. The panel felt that the subject of TA525 and this appraisal were sufficiently similar that a meaningful expectation of consistency applied. Indeed the committee papers seemed to support this view, as they referred to past appraisals including TA525, and so expressly brought those points into consideration. Having rightly set out to take account of practice in relevant past appraisals the committee were obliged to do so reasonably.
22. The panel noted that the committee had responded to the company’s concerns about the preferred treatment effect duration of 3 years in the ACD. The FAD considered treatment effect durations of both 3 and 5 years in this appraisal, and did not prefer a duration of 3 years, which represented a change in the committee’s thinking in response to consultation. Nevertheless, this remained materially different from TA525, where a treatment effect duration of 5 years was preferred.
23. The panel accepted that it was reasonable for the committee not to ask for modelling of a shorter treatment effect duration in TA525 given that the most plausible ICER was comfortably within the range considered cost effective. However, the expectation of consistency means that substantial changes from this approach in similar appraisals would need to be clearly justified.
24. The panel did not judge that the committee had given a clear justification for this difference either in the FAD or during the hearing. The panel judged that even with the longer follow-up data available for this appraisal, there remained very substantial uncertainty about the duration of treatment effect. Whilst accepting Dr Smith’s point that longer follow-up data does not necessarily provide evidence of a longer effect, the panel were not convinced that the committee had yet identified anything in the new data to support a material change in approach from the previous appraisal.
25. In summary, whilst the panel judged that a reasonable approach to duration of treatment effect had been taken both in TA525 and in this appraisal (if each were taken in isolation), there was insufficient justification of why this approach had changed between the two appraisals. This was procedurally unfair given the expectation of consistency set out in the methods guide.
26. The appeal panel therefore upheld the appeal on this point.

**Appeal Ground 1a.2:** It is procedurally unfair to have introduced paragraph 3.19 (“The costs of pembrolizumab are likely underestimated in the model”) in the FAD at a very late stage, without explanation or any opportunity to respond

1. Grant Castle, for MSD, said that the notion that the cost of pembrolizumab may be under-estimated in the model was raised at a late stage in the process. This was mentioned for the first time at the second committee meeting, where MSD did not have a satisfactory opportunity to respond. Following this meeting, MSD contacted NICE to raise their concern. The response from NICE agreed that this was “not ideal” but no attempt was made to address the concern. Whilst this possible underestimation of the cost may not have been a key driver of the decision, it is another factor which had an incremental effect on the committee’s thinking.
2. Dr Lindsay Smith, for NICE, stated that this likely under-estimation of the cost of pembrolizumab was not included in the cost-effectiveness estimates in the FAD.
3. Prof Gary McVeigh, for NICE, acknowledged that this issue arose at the second committee meeting. He said that NICE had been contacted the day before the meeting by NHS England to point out that some patients in the trial had been re-treated. At the meeting, when he asked the company how many patients had been re-treated, they were able to reply without hesitation that it was 5%. Prof McVeigh therefore contended that the company must have been aware of this issue prior to the second committee meeting.
4. Kalpana D’Oca, for MSD, acknowledged that the company knew some patients had been re-treated, but said that they had not known this issue was of concern to the appraisal committee. Had they known this, they would have been able to conduct sensitivity analyses to explore its potential impact.
5. Grant Castle, for MSD, said that the Terms of Engagement for this re-appraisal of pembrolizumab did not include this issue. He argued that the fact that this issue is raised in the FAD demonstrates that it did have an impact on the committee’s decision.
6. The appeal panel concluded as follows. The possible under-estimation of the cost of pembrolizumab in the model was raised by NICE for the first time at the second committee meeting. Although the company were aware that some patients in the trial had been re-treated, they were not aware until this point that this might be a factor in the decision-making in this appraisal. This is a key point. It is not enough that a company must be aware of a fact. They must also be aware that it is relevant, either (as will usually be the case) because that is self-evident, or (as needed to be the case here) because that was made clear. The panel judged that they did not have a satisfactory opportunity to address this point before the FAD was published. Whilst this issue was not a key driver of the final decision, the fact that it was mentioned in the FAD means the panel cannot be confident it had no bearing on the decision. It was therefore procedurally unfair for the issue to be introduced so late in the process without an adequate chance to respond.
7. The appeal panel therefore upheld the appeal on this point.

**Appeal Ground 1b: In making the assessment that preceded the recommendation, NICE has exceeded its powers.**

**Appeal Ground 1b.1:** NICE has breached its legal obligations under human rights and equalities laws

1. The appeal panel received legal advice on this appeal point in advance of the hearing from their legal adviser, Stephen Hocking. MSD had a further opportunity to respond to this, and their advice was also available to the panel and was taken into account. The panel were aware that they may take account of advice and submissions but must apply its own mind to the legal questions raised in the appeal to reach its own decision. The panel was aware that it should not give additional weight to the advice prepared by their legal advisor.
2. At the hearing, Grant Castle summarised MSD’s position. He said that there was consensus that Article 2 of the European Convention on Human Rights was engaged, and that this requires NICE to carry out a fair and rational balancing exercise between the needs of patients affected by the decision and the community at large. There was also consensus that the appraisal committee benefits from a “margin of appreciation” in conducting this exercise. He went on to argue that Article 2 requires the committee to take positive measures to preserve life, yet in this case the committee consistently exercised its discretion in the opposite direction. At every stage, the committee applied a more conservative, negative approach to pembrolizumab compared with atezolizumab (a product that may not extend life). He said that the company does not expect NICE to take a more indulgent view of the evidence in this appraisal compared with others but does expect a consistent approach.
3. Mr Castle also outlined the company’s position on Article 14 and the Equalities Act, set out in more details in their appeal letter and response to the panel’s legal advice. He stated that there are obvious differences between this appraisal and TA525 that are manifestly without reasonable foundation. He argued that this must be discriminatory against patients who would benefit from this treatment compared with others.
4. The appraisal committee was given the opportunity to respond to these points at the hearing but did not wish to add anything.
5. The appeal panel concluded that, although Article 2 of the convention was engaged, it did not make a material difference to the approach that would otherwise have to be taken to this decision by NICE. The panel judged that all of NICE’s processes are directed towards carrying out a fair and rational balancing exercise between the needs of patients affected by the decision and the community at large. It did not accept the company’s argument that life-extending products should necessarily always be given “the benefit of the doubt” (although some NICE processes like the End of Life criteria do indeed do this) or should be subject to special processes or considerations over and above those provided for in NICE’s published procedures. The panel judged that standard NICE processes, applied fairly and reasonably, constitute the necessary positive measures to seek to preserve life. The panel did not agree with the company that in this appraisal the committee had always taken the most conservative approach by default. The committee had considered a range of approaches throughout the appraisal and weighed up various sources of evidence to reach its decision.
6. The appeal panel also understood that Article 14 of the convention requires that the enjoyment of the other rights shall be secured without discrimination on any ground. The panel accepted that patients with locally advanced or metastatic urothelial cancer could have a distinct “status” capable of protection under Article 14. Although the appeal panel has judged that specific aspects of this appraisal were indeed unfair or unreasonable, the panel concluded that it did not add anything to see this as discriminatory. Any unfavourable treatment has arisen as a consequence of the otherwise unfair and/or unreasonable decision, rather than as a consequence of or in connection with any particular status. Therefore it would not be correct to hold that discrimination was a cause of any illegality or that the decision was discriminatory.
7. With regard to the public sector equality duty, again the panel judged that standard NICE processes, applied fairly and reasonably, show due regard to the elimination of discrimination and the advancement of equality of opportunity in most cases. Opportunities to raise equality issues are given during the appraisal process and the FAD specifically notes that no equality issues were identified. There may be particular circumstances where NICE does need to adjust its processes to take account of the needs of a group with a protected characteristic, but that did not apply here. The panel concluded that the fact that a group of patients disproportionately includes older people and people with disabilities is not in itself a reason to deviate from standard processes (as this probably applies in the majority of NICE appraisals). In this particular appraisal, the appeal panel could not identify any specific equality issues that arose during the appraisal and the patient group were not unusually defined by a protected characteristic other than, of course, the disability caused by their disease itself, which the committee plainly took into account.
8. The panel also considered MSD’s argument concerning section 29(6) of the Equality Act. They contend that indirect discrimination could arise from inconsistency in decision-making that has not been properly justified and which has been applied without taking account of the disproportionate impact on people with protected characteristics. While the panel agreed that such a scenario could at the least bear investigation as to whether it was compliant with the Equality Act it was not persuaded that it could easily fit within the concept of indirect discrimination. However it was unnecessary to reach a view on this point because although the panel have agreed under Appeal Point 1.1a that this decision was indeed unfair on the grounds of consistency, the panel did not accept that this constituted discrimination of any form, for the reason given above when discussing Article 14.
9. The appeal panel therefore dismissed the appeal on this point.

**Appeal Ground 2: The recommendation is unreasonable in the light of the evidence submitted to NICE.**

**Appeal point Ground 2.1:** The Appraisal Committee’s assessment of duration of treatment effect and its effect on cost-effectiveness is illogical and unreasonable

1. The discussion on this appeal point at the hearing had some overlap with the discussion under appeal point 1a.1, so this section should be read in conjunction with that section of the decision letter.
2. Grant Castle, for MSD, building on the points raised in the appeal letter, said that MSD had collected unprecedented follow-up data on pembrolizumab which suggested a duration of treatment effect of 5 years or more. He argued that the latest data from KEYNOTE-045 showed clear evidence of benefit at 4 years, and that it was therefore inconceivable that the effect of treatment stops suddenly at 3 years. He argued that a 5-year treatment effect duration is the most plausible. Therefore, the approach taken by the committee in considering treatment durations of both 3 and 5 years simply did not add up.
3. Mr Castle said that the aim of having pembrolizumab in the Cancer Drugs Fund was to allow new data to inform subsequent decisions. He argued that it was unreasonable for the committee to expect “strong evidence” of a five-year treatment effect because the statistical power of the KEYNOTE-045 trial meant that this was not achievable.
4. Dr Lindsay Smith, for NICE, said that the committee had considered the external validity of a duration of treatment effect of 3-5 years in its decision-making. This is shown in slide 16 from the second committee meeting where treatment effect durations used in other appraisals were noted. Dr Smith also pointed out other appraisals of pembrolizumab (not listed in this slide) which have used treatment effect durations of 3-5 years (e.g. TA531 and TA557).
5. Dr Smith was asked by the panel whether he agreed that the other appraisals considered at the second committee meeting suggested that 5 years was a more typical estimate of treatment effect duration in similar appraisals. He disagreed and said that the slide simply demonstrated that a wide range of estimates had been used.
6. Dr Smith, for NICE, emphasised that the committee did not express a preference for a treatment effect duration of 3 years in the FAD. At the ACD stage, the committee had preferred 3 years based on ERG advice, but they modified their position following consultation. At the time of the FAD, the committee concluded that both 3 and 5 years were plausible. However, it did not accept the company’s assertion that 5 years was most plausible. Dr Smith said that the onus would be on the company to provide evidence for this claim and they had not done so.
7. Dr Smith said that the committee recognised that there may be a small group of treated patients with very long survival, and that this was reflected in the modelling, but that there may also be a cohort of patients with long survival in the Standard of Care group. The crucial question is the relative benefit shown by the hazard ratio. The committee had concluded that there was evidence of waning of the treatment effect even before stopping treatment. They had concluded that there was no convincing evidence from clinical trials of benefit beyond 2 years, but they had been persuaded by the evidence of clinical experts to take a less conservative view.
8. Prof Gary McVeigh, for NICE, discussed the argument from the company in their appeal letter that the time varying hazard ratios from KEYNOTE-045 support a treatment effect beyond 3 years. The committee took the view that this data was not reliable because of a small number of events after 3 years.
9. Prof McVeigh stated that some patients in KEYNOTE-045 were re-treated after the 2-year stopping rule (unlike in NHS practice). The company adjusted for treatment switching in the Standard of Care arm but not for re-treatment in the pembrolizumab arm. This re-treatment may have played a particular part in effects seen after longer follow-up.
10. Prof McVeigh said that the ERG concluded that there was no evidence of benefit beyond 3 years because the confidence intervals around the hazard ratio crossed one very early. He therefore argued that the committee had been lenient in deciding to consider a 5-year treatment effect duration.
11. Benjamin Bates, for MSD, said that the hazard ratio confidence intervals are wide because of the small number of patients at risk, not because of an absence of ongoing treatment effect. It would have been unfeasible for KEYNOTE-045 to be powered to detect a treatment effect at >3 years. He argued that the additional data shows that the time-varying hazard ratio seems to plateau, rather than moving towards 1. He acknowledged that there may be waning of the treatment effect and said that this is reflected in the modelling. However, he argued that it is just not plausible that there would be no ongoing treatment effect after 3 years. The company concluded that a 5-year duration of treatment effect was most plausible based on a combination of the data and clinical opinion.
12. Daniel Gallacher, for NICE, explained why the ERG judged a 3-year treatment effect to be the most plausible. He highlighted again that the treatment effect begins to wane as early as 30 weeks, and the confidence intervals for the hazard ratio cross 1 from 60 weeks. This led the ERG to conclude that only a 2-year treatment effect could be robustly supported. They acknowledged that this was probably too pessimistic, so used a 3-year effect in their preferred model. He said that the apparent plateau in the time varying hazard ration could not be relied upon because of the low number of events.
13. Dr Simon Crabb, for MSD, strongly rejected the idea that a 3-year treatment effect duration could be clinically plausible. He said that in his clinical opinion, 5 years is the minimum that is plausible.
14. As discussed under Appeal Point 1a.1, the appeal panel concluded that there is a biologically plausible rationale for thinking that the effect of pembrolizumab may continue after treatment is stopped, but that there is a great deal of uncertainty about the duration of treatment effect.
15. The issue of consistency with previous appraisals has been considered above under Appeal Point 1a.1, so is not discussed further here. In reaching a conclusion on Appeal Point 2.1, the panel has considered whether the assessment of duration of treatment effect was reasonable in the light of the evidence submitted in this appraisal only (leaving aside the issue of consistency).
16. The appeal panel judged that the committee had given a rationale for concluding that a treatment effect duration of 3-5 years could be plausible which if it was looked at in isolation was reasonable. The committee reached this judgement based on the trial data, alongside expert clinical opinion. The committee acknowledged the likely “tail of treatment” effect and took account of this in their decision-making. The panel did not agree with the company’s view that the committee “considered that the most conservative possible estimate must apply by default”. That allegation suggested the committee felt bound to be as conservative as possible, whereas the panel felt the evidence showed the committee exercising judgement, being persuaded on some points, and not always taking the most cautious view. In fact, the ERG judged a treatment effect duration of 2 years to have the most robust evidence, but this was not considered by the committee. The committee were responsive to points raised at consultation, in changing their preferred treatment effect duration from 3 years to at least allowing the possibility of 3-5 years. The committee gave clear reasoning for not placing significant weight on the apparent plateau in the time-varying hazard ration in KEYNOTE-045. Given the substantial uncertainty in the data, it was reasonable to consider a range of treatment effect durations rather than a point estimate. It was also clear that the committee had considered the external validity of their duration of treatment effect at the second committee meeting. The panel did not agree with the company’s position that there had been an unreasonable requirement for “strong evidence” of a 5-year treatment effect. The committee had indeed considered a 5-year treatment effect, but had reached a judgement about how much weight to place on this that took into account the strength of the evidence. The panel did not consider this unreasonable.
17. The appeal panel noted the different views taken by the company and the committee on the most appropriate duration of treatment effect in the appraisal (in the face of complex and uncertain evidence), but felt that both these views could be reasonably held. The appeal panel was satisfied that the committee’s reasoning for reaching its position looked at in isolation was rational.
18. The appeal panel therefore dismissed the appeal on this point.

**Appeal point Ground 2.2:** The Appraisal Committee’s analysis of evidence from the clinical expert in paragraph 3.15 of the FAD is internally inconsistent and its conclusions are unreasonable

1. At the scrutiny stage it was decided to consider this point as an aspect of Appeal Point 2.1. However, at the hearing they were largely considered separately, so this letter also considers Appeal Point 2.2 separately.
2. Grant Castle, for MSD, said that paragraph 3.15 of the FAD mischaracterised evidence given by Dr Simon Crabb, to the detriment of pembrolizumab. The appeal letter said that the statements in paragraph 3.15 that “the clinical expert found it plausible that 5% to 10% of people having pembrolizumab might survive to 10 years after starting treatment” and that “no more than 5% of people treated with pembrolizumab might be alive after 10 years” were inconsistent.
3. Dr Lindsay Smith, for NICE, said that the committee had considered evidence from clinical experts, patient experts and NHS England to externally validate trial data. Ideally the committee would have access to 2 independent experts, but in this case there was only 1 expert (Dr Crabb) who was put forward by the company. In reaching the conclusion that “no more than 5% of people treated with pembrolizumab might be alive after 10 years” they had considered evidence from a patient expert and the Cancer Drugs Fund clinical lead as well as from Dr Crabb.
4. Dr Simon Crabb could not recall exactly what he had said at the meeting in question and felt that this issue was “getting bogged down in detail”.
5. The appeal panel concluded that paragraph 3.15 of the FAD was not internally inconsistent and that the committee’s conclusion on this point was not unreasonable.
6. The appeal panel therefore dismissed the appeal on this point.

**Appeal point Ground 2.3:** The Appraisal Committee’s decisions that: (i) a range of possible ICERs from £48,518 to £70,520 applies; and (ii) the “most plausible” ICER for pembrolizumab is likely to be above £50,000, are unreasonable in light of the evidence submitted

1. Grant Castle, for MSD, said that the committee’s range of possible ICERs from £48,518 to £70,520 was unreasonable because it was based on excessively conservative assumptions. He stated that even if this range were correct, it would be unreasonable to assume the most plausible ICER would be >£50,000.
2. Dr Lindsay Smith, for NICE, said that it is usual for committees to consider a range of ICERs based on their preferred assumptions. In this case, there were three key drivers of the ICERs: duration of treatment effect, adjustment for treatment switching, and the choice of overall survival curve. This generated 12 possible scenarios, and in 11 of these the ICER was >£50,000. The committee had four options for the overall survival curve, and the most pessimistic was not considered in these 12 scenarios. With anything other than the most optimistic curve, even a 5-year treatment effect duration would produce an ICER > £50,000.
3. Dr Smith also explained how the End of Life criteria affect cost-effectiveness thresholds. Without the End of Life criteria, NICE considers treatments to be cost-effective if the ICER is less than £20,000-£30,000. The higher end of this range is used when there is less uncertainty in the modelling, a highly innovative therapy, or important quality of life benefits that are not captured in the model. Dr Smith said that none of these apply here, so the starting point would be closer to £20,000. A weighting of up to 1.7 times is then applied when the End of Life criteria are met. In this case, the lack of robustness in the evidence would likely have meant that a weighting of less than 1.7 was used. So, it is likely that the threshold for pembrolizumab to be considered cost-effective in this appraisal, with End of Life Criteria, would be closer to £34,000 than £50,000.
4. Dr Smith, responding to questions from the panel, acknowledged that this had not been discussed in detail during the appraisal. This was because the committee judged the most plausible ICER to be >£50,000 so further consideration of exactly how the End of Life criteria should be applied was not relevant.
5. Benjamin Bates, for MSD, responded to Dr Smith’s comment on 11 out of 12 ICERs being >£50,000. He said that was not reasonable to consider alternative overall survival curves. In addition, he said that 3 of these ICERs would be very close to £50,000.
6. Prof McVeigh, for NICE, said that in the company’s base case the ICER is verging on the maximum threshold for cost-effectiveness. To accept this ICER, the committee would have to accept that there was no uncertainty about a treatment effect duration of 5 years or the two-stage adjustment for treatment switching. The committee judged that this was simply not tenable and therefore the most plausible ICER must be >£50,000.
7. The panel concluded that it was reasonable for the committee to generate a range of ICERs based on their preferred assumptions and on other plausible scenarios. Whether or not the key assumptions that this range was based on are reasonable is addressed separately in Appeal Points 2.1-2 (on treatment effect duration) and Appeal Point 2.5 (on the use of the 2-stage model versus unadjusted data). These issues are therefore not considered here.
8. Under this ground, the panel has considered the narrower issue of whether it was reasonable for the panel to conclude that the most plausible ICER was likely to be above £50,000.
9. The panel did not rely on Dr Smith’s exposition of the weighting to be given to pembrolizumab being less than 1.7, because this is not discussed in the FAD and the panel were not taken to any contemporaneous document in which the issue was raised. The panel accepted that Dr Smith had this consideration active in his own mind, but was not prepared to accept that it formed part of the committee’s overall reasoning.
10. The panel was persuaded by the committee’s argument that any deviation from the company’s base case (whether in treatment effect duration, overall survival curve, adjustment for treatment switching or indeed other factors) would push the ICER beyond the upper end of the range that can be considered cost effective even if the maximum weighting allowed by the End of Life criteria were applied.
11. Further, if it was suggested by MSD that it was enough that one plausible ICER fell below the usual upper level for recommendation under the end of life criteria, the panel disagreed. An appraisal will typically generate a range of ICERs of greater or lesser plausibility depending on the inputs and assumptions used, and a committee should use that range, and its assessment of the plausibility of the values within it, to reach a holistic judgement whether to recommend or not recommend a treatment. How it does so is very much a matter for its judgement in each appraisal and no attempt should be made to lay down prescriptive general rules.
12. The appeal panel therefore dismissed the appeal on this aspect of point 2.3 but refers also to its decision on Appeal Points 2.1-2 and 2.5.

**Appeal point Ground 2.4:** The statement that “the costs of pembrolizumab are likely underestimated in the model” lacks meaningful explanation and evidence and it is unreasonable to have taken it into account for the purposes of assessing cost-effectiveness

1. At the hearing, the discussion of this Appeal Point overlapped with the discussion of Appeal Point 1a.2, so this section should be read in conjunction with that section of the decision letter.
2. The appeal letter from the company states that the number of patients who were re-treated in the KEYNOTE-045 trial was very small. Therefore, without further analysis, it is impossible to conclude that MSD has materially under-estimated the cost of its product.
3. Grant Castle, for MSD, said that it was unreasonable to consider re-treatment because this would not be done in NHS practice.
4. Prof McVeigh, for NICE, stated that 11 patients in the pembrolizumab arm of the trial were re-treated. The benefits of this were captured in the model, but the costs were not. Prof McVeigh said it was misleading for the company to trivialise this as a small number. Only 24 people were at risk of progression, so a sizeable proportion of those eligible may have been re-treated.
5. The appeal panel concluded that some patients in the trial were re-treated, albeit a relatively small number, and that the benefits but not the costs of this were included in the model. It is reasonable for appraisal committees to consider factors not included in the model that might affect their decision making, particularly when considering how much uncertainty there is about the estimated ICER or range of ICERs.
6. The appeal panel therefore dismissed the appeal on this point.

**Appeal point Ground 2.5:** The conclusion that new data from KEYNOTE-045 “shows the 2-stage method may not be appropriate, and the unadjusted method should also be taken into account” results from a flawed and unreasonable interpretation of the evidence

1. Daniel Gallacher, for NICE, explained that the 2-stage method is used to adjust for the fact that some patients in the Standard of Care arm of a trial may switch to active treatment when their disease progresses. It works by comparing the survival time of those who switched when their disease progressed to the survival time to those who did not switch when their disease progressed. This estimates an “acceleration factor” which is then applied to those who did switch.
2. Mr Gallacher, asked by the panel whether this is usually considered an appropriate method, said that it is. It is one of a range of methods that can be used, and one of the more commonly used because it is relatively simple.
3. Grant Castle, for MSD, said that the 2-stage method is uncontroversial and well-established. NICE agreed to this method when pembrolizumab entered the Cancer Drugs Fund. The terms of engagement for this review suggested that this was a settled point. In this appraisal, he argued that the committee has done an “about turn” by giving equal weight to unadjusted data, without a clear rationale. This was one of the important factors that led the committee to conclude that pembrolizumab could not be recommended.
4. Dr Lindsay Smith, for NICE, emphasised that the committee had not considered the 2-stage model to be unacceptable. Rather, they decided it was one of two models they would like to consider (the other being the unadjusted data). The committee took the view that the 2-stage model was over-optimistic and the unadjusted data was over-pessimistic, so both were useful in reaching their “best guess”.
5. Daniel Gallacher, for NICE, gave a detailed account of the ERG’s concerns about the 2-stage model. Firstly, it assumes a uniform effect (that all patients get the same benefit from switching to five different treatments) which is unlikely to be correct. Secondly, the evidence is that these treatments are not effective in the first 4-5 months (even when given to patients with less advanced disease). The 2-stage model assumes that all patients benefit, but the ERG think this is unlikely to be the case in advanced disease where life expectancy can be short, so patients may die before the treatment effect begins. Thirdly, the ERG did not have detailed information on all potentially important differences between those who switched when their disease progressed and those who did not. This is a “red flag” that the acceleration factor may be capturing the benefit of other prognostic factors rather than the effect of treatment. Fourthly, results were presented without re-censoring, which is not best practice. In addition, there is a wide confidence interval around the acceleration factor.
6. Mr Gallacher, asked by the panel whether the ERG could not have modelled the confidence intervals around the acceleration factor (rather than using unadjusted data) said that this would have been difficult to do.
7. In response to questions from the panel, Mr Gallacher agreed that these issues can result in both “noise” and systematic bias in the estimation of the acceleration factor. He acknowledged that the “true” acceleration factor could be higher as well as lower than the estimate used.
8. Dr Smith was asked by the panel if it was unreasonable to give equal weight to the unadjusted data (which everyone agrees is problematic) and the 2-stage model (which is imperfect but an accepted method). Dr Smith replied that the committee were particularly concerned that two particular issues (the second and third issues described by Mr Gallacher in paragraph102 above) would introduce systematic bias in the acceleration factor and therefore the 2-stage adjusted model was too optimistic.
9. Daniel Gallacher, for NICE, said that his view was that the unadjusted data could be very plausible (if those who switched did not actually receive benefit from treatment, but survived longer because of other prognostic factors).
10. Kalpana D’Oca, for MSD stated that the ERG report acknowledges that the unadjusted data are biased and acknowledges that this bias is likely to be greater than that in the 2-stage model. She also stated that there is still a source of bias in favour of the control arm when the 2-stage method is used (because there are some patients who switched to whom the adjustment is not applied). She commented on the larger magnitude of the acceleration factor in this re-appraisal compared with the original appraisal of pembrolizumab: although the magnitude has increased, it falls within the range of the original confidence intervals, and the confidence intervals have now narrowed so it can be considered more reliable.
11. Ms D’Oca stated that the committee’s support for the unadjusted data seemed to rely on comparison with data from a 2013 study by Bellmunt et al. She argued that this comparison was inappropriate as vinflunine is not established in UK clinical practice.
12. In response to this point, Dr Lindsay Smith, for NICE, said that the Bellmunt study was not relied upon in the final decision, and was not mentioned in the FAD.
13. The appeal panel concluded as follows. There was general acceptance during the hearing that the unadjusted data was biased. At the time of the original appraisal of pembrolizumab in this indication (TA519) the ERG also judged that the unadjusted model was the least appropriate. The panel understood the limitations of the 2-stage model, as set out by Mr Gallacher, and appreciated that these could cause bias as well as imprecision in estimating the acceleration factor. Nevertheless, these same limitations were known at the time of TA519 and the 2-stage model was judged the most appropriate. The committee did not seem to have considered other approaches to dealing with potential biases in the 2-stage model, such as modelling a range of acceleration factors. The panel was not persuaded that it was necessarily unreasonable to give any consideration to the unadjusted data. However, the FAD suggests that the committee gave equal weight to the unadjusted data and the 2-stage model in their decision-making. The panel judged that it was not reasonable to give equal weight to a method that was previously agreed to be the least appropriate, without a clear reason based on new evidence.
14. The appeal panel therefore upheld the appeal on this point.

**Conclusion and effect of the appeal panel’s decision**

1. The appeal panel therefore upholds the appeal on the grounds that
	* Ground 1a.1 – Fundamental differences with the approach taken in TA525 evidence an inconsistent methodology, in breach of the Methods Guide and the principle of procedural fairness
	* Ground 1a.2 – It is procedurally unfair to have introduced paragraph 3.19 (“The costs of pembrolizumab are likely underestimated in the model”) in the FAD at a very late stage, without explanation or any opportunity to respond
	* Ground 2.5 – The conclusion that new data from KEYNOTE-045 “shows the 2-stage method may not be appropriate, and the unadjusted method should also be taken into account” results from a flawed and unreasonable interpretation of the evidence
2. The appeal is dismissed on all other grounds.
3. The appraisal is remitted to the appraisal committee who must now take all reasonable steps to address these points. In particular, the committee should take into account the expectation of consistency with previous appraisals and whether this should influence their approach to estimating the treatment effect duration. To be clear, the panel does not consider that there is a tension between its conclusions on grounds 1a.1 and 2.1-2. Notwithstanding that the Committee’s approach can be regarded as reasonable looked at in isolation, the fact that different approaches have been taken in past relevantly similar appraisals means that, if a different approach is now to be taken, that departure from past approach or practice must be explained, and not only the reasons for the decision in isolation. This requires additional and specific discussion, at least in this case. The panel considers that the committee might be assisted by considering how its processes and its guidance would be seen by patient groups or clinicians, or other stakeholders, and what degree of additional reasoning they would expect to see so that they can understand and have confidence in the Institute’s work.
4. If the Committee conclude that a different approach is needed from that taken in previous appraisals this should be clearly justified. The company should be given a chance to respond to the committee’s concern that the costs of pembrolizumab may have been under-estimated in the model. The company’s response should be considered in reaching a final decision. The committee should re-consider how potential biases in the 2-stage method could be addressed and re-consider the relative weight given to the 2-stage method versus other models in their decision-making.
5. There is no possibility of further appeal against this decision of the appeal panel. However, this decision and NICE’s decision to issue the final guidance may be challenged by applying to the High Court for permission to apply for a judicial review. Any such application must be made within three months of NICE publishing the final guidance.