NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

SINGLE TECHNOLOGY APPRAISAL

APPEAL HEARING

# Advice on cabozantinib for previously treated differentiated thyroid cancer unsuitable for or refractory to radioactive iodine [ID4046].

**Decision of the panel**

## Introduction

1. An appeal panel was convened on 19 September 2023 to consider an appeal against NICE’s final draft guidance (FDG), to the NHS, on cabozantinib for previously treated differentiated thyroid cancer unsuitable for or refractory to radioactive iodine [ID4046].
2. The appeal panel consisted of:

|  |  |
| --- | --- |
| * Dr Biba Stanton | Chair |
| * Professor Peter Groves | Health Service representative |
| * Rachel Russell | Industry representative |
| * Catherine White | Lay representative |
| * Professor Gary Ford | Non-executive director of NICE |

1. None of the members of the appeal panel had any competing interest to declare.
2. The panel considered the appeal submitted by Ipsen Limited (Ipsen).
3. Ipsen was represented by:

|  |  |
| --- | --- |
| * Mark Harries | Head of Market Access, UK and Ireland |
| * Katie Hill | Market Access & communication & external affairs |
| * Lynne Millar | Senior Legal Director, UK and Ireland |
| * Megan Lewis | Associate Director, HEOR and Access, FIECON |

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

|  |  |
| --- | --- |
| * Dr Stephen Smith | Vice chair of Technology Appraisal committee D (chairing the cabozantinib meetings) |
| * Dr Jacoline Bouvy | Programme Director, NICE |
| * Jasdeep Hayre | Associate Director, NICE |
| * Professor Paul Tappenden | Lead of the Evidence Assessment Group for cabozantinib |
| * Rachel Ramsden | Technical Analyst, NICE |

1. The appeal panel’s legal adviser, Alistair Robertson (DAC Beachcroft LLP), was also present.
2. The appeal panel is grateful to Julia Priestley, Chief Executive Officer of the British Thyroid Foundation, who joined the beginning of the hearing and gave moving evidence of the experience of those living with this disease.
3. Under NICE’s appeal procedures, members of the public are admitted to observe appeal hearings and several members of the public and NICE staff observed the proceedings which were held 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:

(a) Failed to act fairly; and/or

(b) Exceeded its powers.

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

1. Dr Mark Chakravarty, NICE Lead non-executive director for appeals, in preliminary correspondence had confirmed that Ipsen had valid grounds for appeal under Ground Two.
2. The appraisal that is the subject of the current appeal provided advice to the NHS on the use of cabozantinib for previously treated differentiated thyroid cancer unsuitable for or refractory to radioactive iodine.
3. Before the appeal panel inquired into the appeal points the following made preliminary statements: Mark Harries (Ipsen) on behalf of the appellant and Dr Stephen Smith on behalf of the appraisal committee. Julia Priestley provided a patient perspective on behalf of the British Thyroid Foundation.
4. The numbering of appeal points in this letter reflects those that were used during the hearing. The text of this document does not represent a verbatim account of the proceedings nor a documentation of the order of events that took place, but rather provides a brief summary of the submissions from both Ipsen and the committee for the points that were discussed relevant to the decisions of the panel.

## Appeal by Ipsen

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

### Appeal point 2.1: That the committee has failed to take a balanced view of the strengths and weaknesses of the survival extrapolation methodologies in the modelled population and that of expert opinion alongside it.

1. Mark Harries, for Ipsen, explained that in the first committee meeting, Ipsen had proposed the use of more traditional parametric survival analysis to predict long-term survival. In that exponential extrapolation model, long-term survival predictions (especially for the population receiving best supportive care) were not in line with expert clinical estimates. Following further consideration, Ipsen proposed an alternative, novel blended survival curve, in the second committee meeting. Mark Harries explained that Ipsen had made use of clinical expert predictions to produce the blended survival curve. Ipsen considered that the committee had unreasonably failed to adopt that blended survival curve. In particular, Mark Harries argued that the absence of a clinical expert in the second committee meeting was detrimental to the committee's consideration of Ipsen's blended survival curve. In Ipsen's view, a clinical expert would have encouraged greater weight to be placed on the blended analysis.
2. Mark Harries acknowledged that during the second committee hearing Ipsen had been unable to answer questions from the committee about the methodology used to produce the blended curve. He also confirmed that the blended curve presented in the second committee meeting contained an error, in that the early data points on the blended curve did not fit observed data from the pivotal COSMIC-311 trial. He explained that Ipsen had subsequently corrected that error, but complained that there had been no opportunity to provide the committee with the corrected model.
3. Megan Lewis, for Ipsen, provided further interpretation of the blended survival curve, explaining that in Ipsen's view, despite the error identified, the model displays better assimilation between the exponential functions and clinical expert estimates.
4. Dr Stephen Smith, for NICE, explained that in the first committee meeting, Ipsen presented a number of potential approaches to extrapolating overall survival rates, and the External Assessment Group (EAG) presented several other potential approaches. He explained that each option takes the form of a curve produced by mathematical functions, which produce predictions for overall survival. He explained that the optimal approach is to validate the modelled curve with actual data. This was not available here owing to an absence of actual data on overall survival, and the committee therefore had to recognise that all the available models were inexact and subject to uncertainty. He noted that the committee were provided with and had considered all the information presented in the meeting well in advance, so they were well acquainted with it by the time of the meeting.
5. Dr Smith explained that Ipsen persuasively explained in the first committee meeting why the exponential extrapolation model should be the preferred model. The committee reached the provisional view in the first meeting that this curve should be preferred, over the EAG's proposal. Professor Paul Tappenden, for NICE, explained further that none of the curves presented in the first committee meeting were perfect, and that this was primarily because the data available from the pivotal trial COSMIC-311 only covered a short period of treatment with cabozantinib. The model should ideally fit the observed data, reflect reasonable beliefs about long term treatment effect and be clinically plausible. In the modelled data after the end of that period, many curves showed an increasing risk of death in the group with cabozantinib over time, and a decreasing risk of death in the placebo group, with the consequence that many of the curves ‘crossed’, showing greater risk in the group with cabozantinib than in the group with placebo. This is not clinically plausible. The exponential extrapolation model for which Ipsen argued in the first committee meeting did not have this problem and was considered to give the least implausible projections of the available options. The committee selected this model, noting that it came with significant uncertainty.
6. Professor Tappenden noted that the EAG and committee had considered that the blended survival curve presented by Ipsen in the second committee meeting must have included an error, but that this had not been confirmed by Ipsen until this appeal hearing. He further noted that although Ipsen had not provided a corrected version of the blended survival model, for it to fit both the observed data and Ipsen’s assertion of improved overall survival, the model would have to show degradation over time and then an increase in survival in the longer term, which would be difficult to explain clinically.
7. Dr Smith explained that despite the model being capable of incorporating observed data, the failure to do so caused the committee to have fundamental concerns in the second committee meeting about the novel methodology deployed. Furthermore, Ipsen had not supplied the underlying formulae to enable scrutiny of that methodology. As a result, the committee did not see good reason to depart from the initial survival extrapolation model presented in the first committee given the lack of evidence and validity available to support the blended survival curve.
8. Finally, Professor Tappenden estimated that had the error been rectified in the blended survival model, it is still likely that the model would produce results that lacked face validity and that the Incremental Cost Effectiveness Ratio (ICER) would have been higher.
9. In relation to clinical experts, Dr Smith explained that clinical experts are only invited to second committee meetings in extraordinary circumstances. Jasdeep Hayre, for NICE, agreed with Dr Smith, and explained that given that the issues remaining in the second committee meeting were economic, statistical issues – the committee did not feel that further clinical input was necessary. Rachel Ramsden, for NICE, added that clinicians did have the opportunity to provide their input in writing during consultation (as shown on page 579 of the committee papers, for example). Professor Tappenden noted considerable divergence in clinical expert views and accordingly doubted that the presence of one clinical expert in the second committee meeting would have significantly influenced the committee's view.
10. The panel further enquired why the error in the blended survival model could not be explained by Ipsen on the day of the second committee meeting. Mark Harries explained that despite Ipsen's initial belief that the model was correct, once the error was identified in the second committee meeting, Ipsen reconsidered and reanalysed to identify the mistake in the application of the methodology. The error has since been corrected.
11. Dr Smith reiterated that had Ipsen returned on the same day as the second committee meeting or shortly thereafter, the rectification of the error could have been taken into account. Nonetheless, he reiterated that the committee did not discard the model – instead, the committee considered both models and decided to prefer the exponential model.
12. In response to questioning from the panel as to apparent similarities between the current appraisal and TA535 Lenvatinib and sorafenib for treating differentiated thyroid cancer after radioactive iodine, Dr Smith explained that in TA535, a different line of treatment was under consideration, using data taken from a different trial.
13. The appeal panel concluded as follows: It reminded itself that this appeal point relates to whether or not the committee had been unreasonable in the view it had taken on the strengths and weaknesses of the survival extrapolation methodologies in the modelled population and that of expert opinion alongside it. Specifically, Ipsen stated that the blended survival curve that they had presented at the second committee meeting had been dismissed by the committee unreasonably and that the overall approach adopted by the committee in this appraisal was inconsistent with that adopted during the previous TA535 appraisal.
14. The appeal panel noted that the committee had identified significant challenges in modelling overall survival after cabozantinib and placebo treatment in this patient cohort in view of the relatively small patient numbers and short follow-up in the COSMIC-311 study. The panel were satisfied that the committee had considered, in a balanced way, a range of models that included parametric and exponential distributions that had been presented by Ipsen and the EAG in an attempt to overcome these challenges, but had identified significant limitations in all of them. The panel noted that the exponential model preferred by the committee was also preferred by Ipsen at the time of the first committee meeting and resulted in the most favourable ICER. The panel judged that the committee had considered the strengths and weaknesses of the blended survival analysis that was presented by Ipsen at the second committee meeting alongside other models.
15. In regard to the blended survival analysis, the panel noted that the committee had concluded that this did not fit the observed data from the COSMIC-311 trial and that there was uncertainty about which function had been used to fit the observed trial data into the survival curve. Ipsen were unable to resolve this at the second committee meeting. During the hearing, it was acknowledged by Ipsen, that an error had been made in generating the blended survival curve that was manifest by differences in this and the EAG exponential survival curve in early follow-up that was most apparent in the best supportive care arm of the study. The panel noted that blended survival curve data points had not been provided by Ipsen which meant that the EAG could not explore this adequately before the second committee meeting. The panel considered that there is a reasonable expectation that Ipsen should be able to describe the methodology that underpinned the generation of the blended survival curve in the meeting and be able to resolve any uncertainties in the minds of the committee, but noted that it had not been able to do so. Consequently, the appeal panel concluded that the committee had acted reasonably in its considerations of the strengths and weaknesses of the blended survival curve and other models. It also considered that it was reasonable for the committee to conclude that, in the face of considerable uncertainty about overall survival which could not be resolved without additional data collection, the exponential function used by the EAG for modelling overall survival in both treatment arms was preferable for its decision-making. The panel considered that the committee's decision-making process around the modelling of overall survival was clear and well described in section 3.7 of the FDG.
16. The appeal panel were satisfied that clinical expert opinion had been sought and considered by the committee prior to the second committee meeting before arriving at its conclusions about overall survival with cabozantinib and placebo treatment. It understood that the attendance of a clinical expert at a second committee meeting is unusual and concluded that in view of the fundamental flaws to the blended survival curve analysis that had been identified, it is unlikely that these could have been adequately resolved at the second committee meeting even if a clinical expert had been present.
17. In regard to the question of consistency between this appraisal and the previous TA535 appraisal, the committee noted that although the patient cohorts in question were similar, there were other significant factual differences between the two appraisals (see appeal point 2.7 also) and considered, under these circumstances, that the committee's conclusions in this case were reasonable.
18. The appeal panel dismissed this appeal point.

### Appeal point 2.2: The committee's decision to selectively use utility values from two different sources for the progression-free survival and progressed disease health states was unreasonable as it is arbitrary, biased, flawed and inconsistent with the NICE Process and Methods Manual (PMG36).

1. Mark Harries, for Ipsen, explained how Ipsen had initially proposed that the committee should assess utility values for both health states (progression-free survival (PFS) and progressed disease (PD)) directly from the COSMIC-311 trial. An alternative would have been to use the values collected from a vignette based study, Fordham et al. (2015).
2. Mark Harries emphasised that it was never Ipsen's intention for the committee to "mix and match" the values from both sources presented. In the COSMIC-311 trial, the PFS state showed a utility value of 0.692 and a PD state of 0.674, which were not as different as Ipsen had expected. In comparison, in the Fordham et al. (2015) study – the health utility values produced were a PFS state of 0.8 and a PD state of 0.5. This larger gap aligned more closely with clinical opinion of the PD state of those suffering from refractory differentiated thyroid cancer (DTC).
3. Mark Harries asserted that combining utility values from different sources is inappropriate and lacks scientific rigour. He noted that NICE's processes do not state that different sources can be used in combination. Mark Harries argued that using different sources creates greater uncertainty.
4. In response, Dr Stephen Smith, for NICE, explained how the committee considers a number of health states, and considers how long patients spend in each state. He explained that the committee needs to ensure that the quality-adjusted life years (QALYs) identified for each health state are as well evidenced as possible. Each health state is considered independently of each other. He explained how the hierarchy of evidence in NICE's Health Technology Evaluations: the Manual (PMG36) (the Manual) is seriously considered by the committee. Trial data is at the top of the hierarchy and in this case, the committee initially anticipated that it would use data from the pivotal COSMIC-311 trial to inform both states. However Ipsen persuasively presented the case that COSMIC-311 data was inaccurate for the PD health state, and the committee heard evidence from experts and others supporting Ipsen's position. Following consideration of that evidence, the committee agreed with Ipsen that the Health Related Quality of Life (HRQoL) for the PD health state reflected in the COSMIC-311 data was not reflective of patient experience. Subsequently, the Committee agreed with Ipsen's suggestion of using the Fordham et al. (2015) vignette study to derive the PD health state.
5. Dr Smith went on to explain that the PFS utility value was stated in Fordham to be 0.8 (compared to a general UK population PFS utility value of 0.82). Given what the committee had heard from patients about the experience of living with DTC, the committee considered the PFS value from the COSMIC-311 trial (of 0.692) to be a more accurate reflection of the severity of the disease. As a result, although the committee were persuaded to depart from the hierarchy in relation to the PD value, there was no reason for the committee to depart in respect of a reasonable PFS value arising directly from trial data in COSMIC-311.
6. Professor Paul Tappenden, for NICE, gave his perspective as lead for the EAG on cabozantinib. He explained that the utility values must reflect the NICE reference case, the target population, and plausibly reflect health state over its full duration. He acknowledged that ideally utility values would be derived from the same source, but said that it is relatively common for that not to be possible.
7. Professor Tappenden went on to explain the four options open to the committee:
8. To use the COSMIC-311 data for both health states, despite the committee's strong belief that the utility value for PD would be biased because of collection difficulties.
9. To use the Fordham et al. (2015) vignette study for both health states, as this would rely on the same source – notwithstanding its low ranking in the evidential hierarchy. He noted that this was the EAG's preferred approach.
10. To use the COSMIC-311 data for the PFS value, and the Fordham data for the PD value. This was the committee's preferred approach, and was a scenario provided by Ipsen in response to a clarification question by the EAG. The EAG saw this to be a reasonable approach.
11. To use data from the DECISION trial (previously used in another appraisal, TA535) as this uses EQ-5D data, despite the fact that it is targeted towards a different patient population at a different line of therapy.
12. Professor Tappenden stressed that none of the above are perfect solutions. He noted that it is not rare for data collection to stop at the point of progression in oncology, and that although using the same source is desirable, there are examples of committees using the EQ-5D from the pivotal trial to inform the PFS value, and external vignette studies to inform the PD value, such as appraisal TA862 (Trastuzumab deruxtecan for treating HER2-positive unresectable or metastatic breast cancer after 1 or more anti-HER2 treatments).
13. Mark Harries advocated for the need for consistency to prevent what he described as innumerable possibilities in terms of combinations.
14. Dr Smith acknowledged that although using the same source is desirable, the main aim is for health economic models to reflect clinical reality as closely as possible. The committee were of the view that adopting the combination of utility values was the best way to achieve this. In explaining the committee's decision making, Dr Smith confirmed that there was no other alternative literature sources for EQ-5D data (consequent of the small patient population), and that mapping alternative data is only usually deployed where no EQ-5D data had been collected (which is not the case given the EQ-5D data used in COSMIC-311). Dr Smith also discounted proxy conditions as an option.
15. The appeal panel concluded as follows: It reminded itself that this appeal point relates to whether or not the use of utility values from two different sources for the PFS and the PD patient groups that informed the calculations of the cost effectiveness of cabozantinib was unreasonable in terms of its logic and scientific validity. In reaching its judgment on this appeal point, the panel noted that the committee were aware that the Manual says that (HR-QoL) data measured by patients is preferred and says that the use of utility values derived from different methods cannot always be compared. Nonetheless, the panel also noted the considerable challenges encountered by the committee in this appraisal bearing in mind the limited HR-QoL data that was available, particularly in the PD arm of the COSMIC-311 study.
16. The appeal panel noted that the committee had considered the option of using utility values from the same sources in the PFS and DP groups but had dismissed this as a potentially misleading approach. The panel noted that this was because, had the utility values in these two groups both been derived from Fordham *et al.*(2015), then they would not have included HR-QoL data derived from patients in the COSMIC-311 pivotal study and the utility value in the PFS arm would have been almost as high as in the general population (which appeared implausible), and that had the utility values both been derived from the COSMIC-311 study, then the utility values in the PD arm would have been very uncertain (because HR- QoL data collection ceased shortly after progression was identified in the study) and the ICERs would have been higher. The appeal panel concluded that these concerns were reasonable, logical, and scientifically valid.
17. The appeal panel noted that the committee had concluded that the use of utility values in the PFS and PD patient groups derived by different methods and sources was preferable and considered that this had been a reasonable conclusion bearing in mind the data available. In reaching this decision, the panel took account of the fact that the use of utility values from the PFS arm of the COSMIC-311 study meant that this was based on the population being appraised and that HR-QoL had been measured directly in study patients. It was persuaded that, in the face of limited utility data in the PD arm of the COSMIC-311 study, the committee had explored all other potential and relevant sources for utility values in this group and that the use of utility values from the vignette study, Fordham *et al.* (2015) was the most credible alternative available to them.
18. The panel noted that the Manual does not explicitly exclude the use of utility values derived from different methods and sources and recalled that they had been informed by NICE during the hearing that this approach has been previously adopted in the appraisal of a treatment for breast cancer. The panel concluded, therefore, that the committee had acted reasonably in their selection of utility values from different sources in the PFS and PD arms; that this was scientifically valid; and that the rationale for this approach was clear and well described in section 3.9 of the FDG.
19. The appeal panel dismissed this appeal point.

### Appeal point 2.6: The committee's conclusions regarding the appropriate ICER threshold for this appraisal do not assess uncertainty in a balanced way nor do they take into account the likelihood of decision error and its consequences in accordance with NICE's Methods Guide.

1. Mark Harries, for Ipsen, referred to paragraph 6.2.28 of the Manual, which states that the committee ought to take into account the likelihood of decision error and its consequences. He argued that this had not occurred in this case. He noted that all of the ICER variations presented by Ipsen at Table 1 of its Appeal Letter are below £30,000 per QALY gained.
2. Dr Stephen Smith, for NICE, explained that the generally acceptable threshold for a positive recommendation for use of a technology in the NHS is £20,000 per QALY. He noted that it is rare for an appraisal committee to be able to be completely certain about all the factors driving its ICER calculation. He explained that if the NHS is paying the same amount for a technology where the ICER is relatively more certain, and for a technology where the ICER is relatively less certain, then the risk to the NHS is greater in respect of the technology where the ICER is less certain. He noted that a committee may recommend a technology with an ICER above £20,000, and potentially up to £30,000, where the data underlying the ICER is very certain. In this case, he described the data as of 'maximum uncertainty on every measure' and highlighted the particular uncertainty in relation to overall survival and health state utilities, as discussed in appeal points 2.1 and 2.2.
3. In light of the uncertainty, at the first committee meeting the committee considered that an appropriate ICER would be £20,000 or less. Following consideration of Ipsen's evidence on the small population and the unmet need, by the second committee meeting the committee considered an ICER up to around £25,000 would be acceptable.
4. Dr Smith said that the committee considered both the likelihood of decision error and the risk (or hazard, or consequence) of such error arising, throughout its decision-making. He said the committee applies extra consideration where, as here, uncertainty is greater. He said that where the committee is presented with a range of scenarios, a few of which are cost effective and many of which are not, that will lead the committee to think that the likelihood of decision error (in the event of a positive recommendation) is greater.
5. Dr Smith explained that the risk arising from decision-error grows with the likelihood of decision-error. The higher the likelihood, the greater the risk.
6. The panel questioned whether the committee believed that their decision making had been accurately reflected in the drafting of the FDG. Dr Smith described the difficulty in striking a balance between brevity and thorough explanation, and acknowledged that a fuller explanation could have been provided in the FDG.
7. Jasdeep Hayre, for NICE, added that assessing decision error is difficult, as it is not defined in the Manual. Although only approximately 72 patients per annum would benefit from the treatment currently, the figure is likely to increase in the future – therefore the argument in respect of a small population affected is heavily subject to change. Jasdeep Hayre also noted that under the Manual, a severity modifier is applied of x1.2 QALYs gained.
8. The appeal panel concluded as follows: It reminded itself that this appeal point related to whether or not the committee had been reasonable in their assessment of the appropriate ICER threshold and specifically the extent to which they had taken account the likelihood of decision error and its consequences in accordance with the manual.
9. The appeal panel noted that in this appraisal, there was significant uncertainty in the minds of the committee about the available data and assumptions that underpinned the assessment of cost effectiveness of cabozantinib in the patient cohort under consideration. They noted that this was because of the small population; the limited data on overall survival in the treatment groups from the pivotal trial; and concerns about the health utility values used in the cost effectiveness calculations (which are discussed in appeal point 2.2). It was also noted that most of the ICERs were in the £25,000 to £30,000 range. The panel were satisfied that these uncertainties were considered by the committee in arriving at its conclusions about the acceptable ICER threshold.
10. The panel noted that the committee's acceptable ICER threshold was greater than the normally accepted NICE threshold of £20,000 per QALY gained and had widened and increased from the 'lower end' of the £20,000 to £30,000 range in the draft guidance after the first committee meeting to the 'lower half' of the £20,000 to £30,000 range in the FDG after the second committee meeting. They understood that this was in acknowledgment of the unmet need in this patient cohort. The panel also noted, however, that the committee had concluded that the acceptable ICER threshold could not be increased further in view of the significant uncertainties in the data and assumptions that under-pinned the assessment of the cost effectiveness of cabozantinib.
11. The panel were satisfied that in their decision-making around defining the acceptable ICER threshold, the committee had adopted a balanced approach and had taken into account benefits not captured in the QALY calculations. The panel were also persuaded that in defining the acceptable ICER threshold the committee had reasonably taken into account the possibility of an error in giving a positive recommendation to a treatment that was not demonstrably cost effective and that it was aware of the potential financial consequences of this for the NHS and the risk of displacing other cost-effective treatments.
12. The panel also concluded that the committee had reasonably taken into account the possibility of an error in denying a potentially cost-effective treatment to this patient cohort but had been limited in the extent to which it was able to mitigate this because of the uncertainties in the evidence available.
13. The appeal panel concluded that the committee had taken a balanced and reasonable approach in their assessment of the appropriate ICER threshold and the possibility of decision error and its consequences in this appraisal but considered that the description of this in the FDG could have been clearer and more explicit.
14. The appeal panel dismissed this appeal point but would advise the consideration of minor clarification to the wording of the FDG to make the decision-making process around the specific question of decision error and its consequences more explicit.

### Appeal point 2.7: The committee's conclusion regarding the plausible ICER and maximum acceptable ICER thresholds is unreasonable as it is arbitrary and mired in obfuscation.

1. Mark Harries, for Ipsen, described the range of different acceptable ICER thresholds communicated by NICE and received by Ipsen in March, April, June and May 2023. He complained that NICE had 'moved the goalposts' without explaining why, and argued that the differences were inexplicable in light of the fact that none of the committee's preferred assumptions had changed. Mark Harries compared the current recommendation with the positive recommendation in TA535, where the committee applied cost effectiveness thresholds above those normally accepted when recommending Lenvatinib and Sorafenib.
2. Dr Stephen Smith, for NICE, was of the view that the sentences used to communicate the acceptable ICER threshold ranges were identical in the first two instances (being "the low £20,000's per QALY gained" and "lower end" of the £20,000 to £30,000 range) and identical in the second two instances ("lower half" of that range and "in the middle of" that range). Dr Smith described them as two 'pairs' of descriptors, with the first pair describing the lower threshold that the committee considered in the first appraisal meeting, and the second pair describing the higher threshold that the committee settled upon in the second committee meeting.
3. Dr Smith explained that the change between the first pair and the second pair of descriptors was as a result of Ipsen's success in the first committee meeting in persuading the committee of the degree of unmet need. Following questioning from the panel, Mark Harries agreed that the pairs of threshold ICERs were clearly communicated.
4. In relation to the recommendations made in TA535, Dr Smith reiterated that an appraisal committee is not bound by previous appraisal outcomes, and that all appraisals are considered on their own merits.
5. Dr Jacoline Bouvy, for NICE, explained that a direct comparison with the TA535 recommendation cannot be made as there was less uncertainty and a different treatment line considered in that appraisal.
6. Both Dr Smith and Jasdeep Hayre, for NICE, warned against the 'chilling effect' of limiting or inhibiting informal communications between NICE and other companies during appraisals by imposing an unrealistic burden of wording consistency on those communications.
7. The panel enquired whether a referring to a numerical range might be more helpful than a narrative description of the range. Dr Smith responded that this would be unnecessarily inflexible.
8. The panel concluded as follows: It reminded itself that this appeal point relates to whether or not the committee conclusions about the most plausible ICER and the maximum acceptable ICER threshold were reached in a reasonable manner and communicated clearly in the FDG. The panel also considered concerns about consistency with the previous appraisal TA535 under this appeal point.
9. The appeal panel noted that elements of this appeal point have already been referred to and considered in the appeal panel's conclusions to appeal points 2.1, 2.2 and 2.6.
10. The panel noted the deterministic cost-effectiveness estimate generated by the committee's preferred assumptions was £28,200 per QALY gained. The panel were satisfied that the committee's most plausible ICER was arrived at in a reasonable and balanced manner; that the underlying assumptions had been consistent throughout the appraisal having not changed between the first and the second meeting; and that these were communicated clearly in section 3.15 of the FDG.
11. The panel noted that the committee's maximum acceptable ICER threshold was greater than the normally accepted NICE threshold of £20,000 per QALY gained and that the range of the maximum acceptable ICER had widened and increased between the first and second committee meetings in recognition of the unmet need in the patient cohort under consideration (see appeal point 2.6). The panel noted that, during the hearing, Ipsen agreed that this change from the “lower end” to the “lower half” of the ICER range was reasonably clear to them in communications from NICE, but their concern was that they did not understand the reason for this change. The panel noted that this issue is specifically addressed in the response to consultation comments, and that the FDG did refer to the role of unmet need in its decision making. The panel also noted that the significant uncertainties in the data and assumptions that under-pinned the assessment of the cost effectiveness of cabozantinib in this patient cohort had limited the committee’s preparedness to increase the maximum acceptable ICER threshold further, and that this was also referred to in the FDG.
12. Overall, the panel were satisfied that the committee’s conclusions about the ICER threshold were reached in a reasonable and balanced manner and that the change in the maximum acceptable ICER range during the course of the appraisal was intended to provide more flexibility. The panel considered, however, that minor changes might be considered to the wording of section 3.13 of the FDG to make it clearer that the setting of the acceptable ICER threshold at greater than £20,000 per QALY gained was driven by the committee's conclusions about unmet need rather than by uncertainties in the estimates of cost effectiveness.
13. In regard to the issue of consistency with the previous appraisal TA535, the appeal panel noted that there were significant factual differences between that and the appraisal under consideration in this appeal. Most importantly, it noted that the degree of uncertainty in the plausibility of longer-term survival extrapolations was less in the minds of the committee in TA535 than in this appraisal, in view of the evidence available to them, and that this had informed their conclusions about the acceptable ICER threshold. The appeal panel concluded that there is no obligation on the part of the committee to adopt a similar approach or draw similar conclusions to the previous appraisal TA535 (see appeal point 2.1).
14. The appeal panel dismissed this appeal point but would advise the consideration of minor clarifications to the wording of section 3.13 of the FDG to clarify the reasoning behind the setting of the acceptable ICER threshold in the lower half of the £20,000-£30,000 range.

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

1. The appeal panel dismissed the appeal against this appraisal on all grounds but draws the attention of NICE to paragraphs 61 and 75 of this decision letter that recommend the consideration by NICE of minor clarifications to the FDG.
2. 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.