NICE National Institute for Health and Care Excellence

Model Feedback Form

Topic: Virtual chromoendoscopy for real-time assessment of colorectal polyps during colonoscopy

Name: [redacted] **Date:** 10 November 2016

Organisation: Olympus

Issue 1 Technical error in cost calculations and their relationship to diagnostic accuracy inputs

<table>
<tr><th>Description of problem</th><th>Description of proposed amendment</th><th>Result of amended model or expected impact on the result (if applicable)</th><th>EAG response</th></tr>
<tr>
<td>When considering optical diagnosis, the implications outlined on page 142 of the report suggest the comparator with the worst diagnostic accuracy may have the highest overall costs, because the short-term cost ‘savings’ achieved by identifying fewer adenomas and performing fewer polypectomies should be offset by the long-term costs of more missed adenomas and incorrect follow-ups.

Following this logic, in the base case analysis, FICE has lower short-term costs than NBI which is driven by the number of TP and FP which, in turn, is driven by the diagnostic accuracy inputs. As FICE has a better specificity input than NBI, it translates into more FP and, therefore, greater overall costs.

However, if the specificity input for FICE is made lower than that of NBI (i.e. worse) then following the logic</td>
<td>Given the limited model transparency it has not been feasible to diagnose the issue or where in the calculations the bias is occurring. As such, we recommend the whole model functionality and calculations are carefully reviewed to correct for any calculation errors and to ensure that superior diagnostic accuracy is associated with lower overall costs.</td>
<td>It is uncertain what affect a review and any subsequent changes might have on the overall model results but, at minimum, it should result in lower overall costs for superior diagnostic accuracy.</td>
<td>Firstly we were unable to replicate the example results given by the company. When we run the analysis with a sensitivity for FICE of 0.76, the long-term costs for FICE are £308.31, rather than £306.08.

<table>
<tr><td colspan="4">Base case diagnostic accuracy inputs: cost outputs</td></tr>
<tr><td></td><td>NBI</td><td>FICE</td><td>i-Scan</td></tr>
<tr><td>DT costs</td><td>607.46</td><td>603.13</td><td>606.21</td></tr>
<tr><td>LT costs</td><td>308.39</td><td>298.12</td><td>303.53</td></tr>
<tr><td>FM costs</td><td>915.85</td><td>901.25</td><td>909.74</td></tr>
<tr><td colspan="4">Base case diagnostic accuracy inputs for NBI and i-Scan, and</td></tr>
</table>
</td>
</tr>
</table>

outlined on page 142 of the report, the overall costs for FICE should be higher as the long-term consequences of less accurate identification of adenomas (and corresponding costs) should offset any short-term 'savings'. If the specificity input for FICE is reduced beneath that of NBI's, however, the costs for FICE are still lower (see table below for an example), suggesting an intrinsic bias / technical error in the model.

Base case diagnostic accuracy inputs: cost outputs			
	NBI	FICE	i-Scan
DT costs	607.46	603.13	606.21
LT costs	308.39	298.12	303.53
FM costs	915.85	901.25	909.74
Base case diagnostic accuracy inputs for NBI and i-Scan, and FICE sensitivity; updated specificity of 0.76			
DT costs	607.46	606.24	606.21
LT costs	308.39	306.08	303.53
FM costs	915.85	914.55	909.74

FICE sensitivity; updated specificity of 0.76			
DT costs	607.46	606.24	606.21
LT costs	308.39	308.31	303.53
FM costs	915.85	914.55	909.74

DT=decision tree; FM=full model; LT=long-term

Secondly, the costs are not only influenced by the specificity values but also by the sensitivity. The cost for FICE is lower than for NBI because the differences in sensitivity between i-scan and FICE.

DT=decision tree; FM=full model; LT=long-term			

Issue 2 Lack of transparency in short-term cost calculations

Description of problem	Description of proposed amendment	Result of amended model or expected impact on the result (if applicable)	EAG response
Published literature consistently demonstrates DISCARD strategies with virtual chromo-endoscopy (VCE) technologies to offer cost savings over a 'histopathology' strategy based on two components: Firstly, the avoidance of histopathological exams based on the feasibility to diagnose optically and correctly predict surveillance intervals combined with minimal risk for long-term outcomes (resect & discard strategy). Secondly, savings by avoiding unnecessary resections of hyperplastic polyps and associated adverse events (leave strategy). Following this logic, the comparator with superior diagnostic performance should achieve short-term cost savings by the correct identification of more adenomas / hyperplastic polyps which avoids the cost of unnecessary resection / histopathology testing of hyperplastic polyps Carrying this logic through to the VCE comparisons, the VCE comparator with superior diagnostic performance should:	Firstly, a thorough review of all cost calculations in the model should be conducted to ensure they are accurate, with greater transparency provided in the model to account for why superior diagnostic accuracy inputs do not lead to lower costs. We recommend the calculation of the short-term costs be particularly reviewed to correct for any calculation or interpretation errors and to ensure that implementation of the optical diagnosis (including 'leave in situ' and 'resect and discard') is associated with lower costs. Explicit guidance should be provided in the report to aide interpretation of the model results. Secondly, comparator specific low confidence prediction rates should be incorporated into the model to better reflect differences across comparators (c/f DAP32 DAR Comments Table.doc) and ensure that the 'logic' of the 'resect and discard' and 'leave in situ' strategies is accurately captured.	It is uncertain what affect a review and any subsequent changes might have on the overall model results but, at minimum, it should result in lower short-term and overall costs for superior diagnostic accuracy inputs.	We disagree that step 3 should outweigh step 1. The higher costs due to correctly identifying more adenomas and the lower costs due to identifying fewer hyperplastic polyps depends on the differences for sensitivity and specificity for i-scan and FICE. In this case, the additional costs due to identifying more adenomas is greater than the reduction in costs due to identifying fewer hyperplastic polyps and hence FICE is cheaper than i-scan. This is due to the larger difference in sensitivity (0.81 vs 0.96) than for specificity (0.85 vs 0.91) for FICE and i-scan.

1. Correctly identify more adenomas o Outcome: higher cost due to increased polypectomies performed and associated costs 2. Correctly identify more hyperplastic polyps with high confidence o Outcome: no cost (leave in situ) 3. Identify fewer hyperplastic polyps with low confidence o Outcome: lower cost as fewer resections / histopathological testing needed Collectively, the cost savings from step 3 should outweigh the additional costs of step 1, resulting in the product with superior diagnostic accuracy inputs having the lowest short-term costs. Given the diagnostic accuracy inputs used in the base case analysis, i-Scan should, therefore, have lowest short-term costs. However, in the base case analysis, despite having worse diagnostic accuracy, FICE is associated with lower costs than i-Scan. A potential misinterpretation of the above calculation may exist in the implementation of the model. While the model appears to appropriately capture the costs of step 1			

above, by applying the same low confidence prediction rate across predictions the cost savings of step 3 are not accurately captured and, therefore, the model is biased against the comparators with better diagnostic accuracy inputs.			

Issue 3 Inconsistency in costs per procedure

Description of problem	Description of proposed amendment	Result of amended model or expected impact on the result (if applicable)	EAG response
There is inconsistency in the equipment costs reported in the report for the different comparators and how they translate into the scenario 7 analysis – on page 29-30 of the report, the endoscopy system costs are very similar for NBI and i-Scan (i.e. £87,385 vs. £83,616, respectively) but the total cost per endoscopy calculations in Table 73 of Appendix 11 show a big discrepancy between NBI and i-Scan (with NBI costs being considerably higher: £232.85 vs. £160.64, respectively). Since both system costs are based on the same calculation – i.e. the whole endoscopy system including processor, endoscope and annual maintenance – it is unlikely the cost per procedure between these comparators would show such a large discrepancy. Furthermore, as the i-Scan data are commercial in	The same or similar total cost per endoscopy should be applied to both NBI and i-Scan in the model and greater transparency should be provided as to how these costs were calculated. This primarily comes from the following rationale: - The cost of the system, as described in the left column looks similar between technologies (p29-30 of report) - If a discounted price is set or average sales price (ASP) was used, the same figures should be applied for all technologies – however, the total system costs described in the report (p29-30) appear to be list prices - The cost of the endotherapy devices should be calculated as the average value from major manufacturer costs because the endoscopist has the option of several products to choose per exam which are produced by different manufacturers, such as Olympus, Boston Scientific, Cook and	The use of a similar cost per endoscopy for NBI and i-Scan is expected to ensure a more accurate calculation of overall costs in the model.	The scope/ system/ maintenance costs in our Scenario Analysis 7 were annuitized and adjusted for throughput. The method we used has been described in Appendix 11 of our report. Although the system costs are similar between NBI and i-scan, [redacted] It should also be noted that for our base case analysis we have not included acquisition costs for VCE. The costs have only been included for scenario analysis 7.

confidence, it is not feasible to deduce from the model itself how the analysis 7 table on the Scenario Analysis worksheet values are derived.	several smaller companies. The same cost should be applied across the comparators since all products can be used with any scope, regardless of the manufacturer (i.e. no compatibility issues) provided the working channel diameter fits. With an estimated market share of 55% for Olympus, 15% for Boston Scientific, 15% for Cook Medical, and 10% for Diagmed Healthcare, the average price (list price) per unit of snares should be calculated. Likewise, an average list price per unit for the forceps should be applied assuming a 10% market share for Olympus, 50% for Boston Scientific, 20% for Cook Medical, and 15% for Diagmed Healthcare. It is recommended that these average values should be calculated for and applied to all comparators.		

Issue 4 Inappropriate use of confidence level data

Description of problem	Description of proposed amendment	Result of amended model or expected impact on the result (if applicable)	EAG response
Inconsistency in definitions of high confidence across technologies and studies suggests the use of a single low confidence value for all model comparators is inappropriate. Furthermore, there is a lack of transparency over the use of how this value and calculations of 1 minus this value to generate high confidence data interrelates	Use comparators-specific low confidence model inputs to better reflect differences across technologies (c/f DAP32 DAR Comments Table.doc). Provide better description in model report about how the use of the low confidence value and 1 minus the low confidence value that is used to generate high confidence data in the model interrelates with the use of high confidence sensitivity and specificity inputs.	Given the low confidence value applied in the model appears to influence the number of polypectomies performed for each comparator which, in turn, drives the associated costs, differential values for each comparator will influence the overall model results and increase / decrease polypectomy-related costs	We assumed that all technologies would have the same value for the proportion of characterisations made with low confidence, due to the limited data from the i-scan and FICE studies. For i-scan there was only one study, which had a proportion of low confidence of 20%, similar to the value used from the NBI

with the use of high confidence sensitivity and specificity inputs – as such, it is unclear if 'double counting' is occurring for high confidence predictions.		depending on how the new values deviate from the original input.	studies (21%). For FICE, there were no data available. There is no possibility for double counting in the model of people with high or low confidence assessments. People who are assessed at low confidence receive histopathology, which is assumed 100% accurate, and do not undergo real-time assessments using virtual chromoendoscopy; there is no overlap between the groups, and therefore no potential for double counting.

Please add further tables if necessary.