NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

CHTE methods review

Exploring uncertainty

Task and finish group report

August 2020

Contents

Executive summary

1. A task and finish group was organised to consider current Centre for Health Technology Evaluation (CHTE) methods for **exploring uncertainty**. The group considered approaches used across the Centre, by other Health Technology Appraisal (HTA) bodies, and in key academic literature to identify evidence for potential **cases for change** to improve the way in which decision uncertainty and its component uncertainties are characterised, explored and presented to decision-making committees. The group's considerations were also informed by previous NICE work, methods guidance published by the Treasury, a workshop about uncertainty in non-HTA sectors, and a survey of CHTE committee members. Topics considered by the group included key aspects of parameter uncertainty (for example, probabilistic analysis), structural uncertainty (for example, scenario analysis), extrapolation beyond available data, the use of value of information analysis and approaches to presenting uncertainty.

2. A summary of the group's findings and proposed cases for change is provided in the tables 1 to 6.

Table 1 Parameter uncertainties: findings and proposed cases for change

Topic	Current methods	Case for change	Comments
Probabilistic results in decision making	Most programmes in the Centre for Health Technology Evaluation (CHTE) currently ask for probabilistic analyses, but decision making often relies more heavily on deterministic and scenario analyses. These do not reflect parameter uncertainty and may be inaccurate for non-linear models. Appropriate probabilistic analyses are needed for the best estimate of cost effectiveness.	Yes – major	- Starting position is that probabilistic analysis should be done, to get the best estimate of cost effectiveness. If it is not, this should be appropriately justified (for example, for a cost-minimisation analysis in a fast-track technology appraisal). - Committees' preferred cost-effectiveness estimates should be probabilistic. Scenario analyses should also be probabilistic. - This will need greater focus on model parameterisation. - Notable change for the Medical Technologies Evaluation Programme (cost-consequence analyses should be probabilistic).
Model convergence	Probabilistic analyses usually use an arbitrary number of model runs (for example, 1,000). This may not be enough to minimise avoidable Monte Carlo error, which contributes to the overall uncertainty.	Yes – minor	- Companies should show that the number of probabilistic model runs used is enough to minimise Monte Carlo error. - No implementation issues identified.

Topic	Current methods	Case for change	Comments
Ordered parameters	Some parameters are correlated with each other by having a natural order (for example, quality of life with 'mild' disease is logically better than with 'severe' disease). Existing approaches to include this correlation in models may be suboptimal.	Yes – minor	- Consider approaches that impose neither restrictions on distributions nor unsupported perfect correlation. - No implementation issues identified.
Implausible parameters	Uncertain parameters may be included in analyses with unsupported or clinically implausible values. These analyses may be given undue weight in decision making.	Yes – clarification	- Existing methods guide text states that analyses should be plausible. However, this message needs to be clearer. - Inputs values should be consistent with the data and plausible. - Expert elicitation should be used to inform uncertain parameters, rather than just using a wide range of values. - No implementation issues identified.
Probabilistic one-way sensitivity analysis	Deterministic sensitivity analyses are often presented to committees. However, these are less informative when parameters are correlated and do not account for the likelihood	Yes – minor	- Probabilistic one-way sensitivity analysis may be explored. - Not 'must be', as the methodology is new. - No implementation issues identified.

Topic	Current methods	Case for change	Comments
	of each parameter value along the range used.		
Threshold analysis	Threshold analysis may be useful to identify 'switching' values; the value a parameter would have to take to alter the cost-effectiveness conclusion. Only 1 CHTE methods guide includes threshold analysis.	Yes – minor	- Threshold analysis can be used to explore highly uncertain parameters, as it may be informative for committees. - It should not be used to identify cost-effective subgroups, or for parameters that are highly correlated with other parameters. - No implementation issues identified.

Table 2 Structural uncertainty: findings and proposed cases for change

Topic	Current methods	Case for change	Comments
Developing the model concept	Current methods outline that structural modelling assumptions should be clearly documented and supported. However, a transparent account of how a model was developed is not always provided.	Yes – minor	- Model development, including the choice of structure, should be transparently documented and justified. - Details of expert involvement to inform structural assumptions should be provided. - No implementation issues identified.
Scenario analysis	Scenario analyses are presented to committees. Scenarios that use clinically implausible	Yes – minor	- Analyses based on demonstrably implausible assumptions are only informative if they show that the scenario of interest does not

	assumptions may be given undue weight in decision making. It may be possible to incorporate structural uncertainties into a probabilistic analysis, rather than presenting multiple distinct scenario analyses.		matter (that is, does not materially affect conclusions about cost-effectiveness). - Incorporating structural uncertainties into probabilistic analyses can be considered, but the approach used should be carefully considered and reported. - No major implementation issues, although it is recognised that it may be challenging to implement the methods identified in a robust way.

Table 3 Extrapolating beyond the data: findings and proposed cases for change

Topic	Current methods	Case for change	Comments
Treatment benefits over time	Current Centre for Health Technology Evaluation (CHTE) methods list scenarios to explore the effect of long-term treatment benefits, but they include implausible scenarios (that is, that treatment benefit would stop for people who remain on treatment).	Yes – clarification	- The wording should change to reflect more reasonable scenarios to explore. - No implementation issues identified.
Flexible survival models	There is increasing evidence that 'standard' parametric models frequently presented to committees are inferior to more	Yes – minor	- Flexible extrapolation methods should routinely be considered as part of the toolkit of survival models available.

	complex, flexible methods at predicting long-term survival, particularly for immunotherapies.		- Flexible models are already increasingly being accepted by CHTE committees. A Technical Support Document will provide technical guidance on these approaches.
Selecting the appropriate extrapolation curve	CHTE committees often use predicted survival estimates to inform choice of extrapolation. When selecting the most suitable function, comparing how hazards change over time may be more informative than comparing time-to-event estimates. Some functions may predict decreasing confidence intervals (that is, less uncertainty) the further it gets from the observed data.	Yes – minor	- The clinical plausibility of hazard functions should be assessed when comparing alternative models for extrapolation. - The uncertainty in extrapolation over time should be considered; decreasing confidence intervals may understate the true uncertainty. - No major implementation issues identified.
Adjusting for treatment switching	Statistical methods to adjust for treatment switching in trials (crossover) are widely used. The method used may introduce uncertainty.	Yes – minor	- Uncertainty associated with the adjustment method used should be accounted for. - The proposed text is consistent with existing text on uncertainty introduced by using surrogate relationships in models. - No implementation issues identified.

Table 4 Value of information analysis: findings and proposed cases for change

Topic	Current methods	Case for change	Comments
Consistency across CHTE methods guides	The Diagnostics Assessment Programme (DAP) and the Medical Technologies Evaluation Programme (MTEP) guides state that decision-making committees can consider the value of information from evidence generation to inform important evidence gaps, but they do not go into close detail about what methods should be used. Like DAP and MTEP, recommendations that include evidence generation are permitted in the Technology Appraisals (TA) Programme, but the TA methods guide does not mention value of information analysis as a tool for decision makers.	Yes – major	- The task and finish group did not reach consensus about how value of information methods should be captured in the Centre for Health Technology Evaluation, if at all. - A simple approach (Expected Value of Perfect Information; EVPI) may be easy to implement and quantifies the maximum value of the parameter uncertainty. However, it is less informative than more involved methods, it is unclear how it would be used in the decision-making process, and it may itself be subject to uncertain estimation. - The working group agreed to proceed with consultation on the use of EVPI as a useful supplement to probabilistic analysis.

Table 5 Presenting uncertainty: findings and proposed cases for change

Topic	Current methods	Case for change	Comments

Overall uncertainty	A single textual summary of the key sources of uncertainty may be useful as an addition to the Centre for Health Technology Evaluation methods.	Yes – minor	- Review groups should present their overall assessment of the uncertainty, including the relative effect of different types of uncertainty, and whether some are inherent and unresolvable. - This is not currently done by most review groups but is unlikely to be a significant additional burden.
Net benefit ranks	When more than 2 alternative options are being compared, cost-effectiveness acceptability curves become less useful for decision making (becoming cluttered and potentially concealing uncertainty). Net benefits and net benefit rankings may be more informative.	Yes – minor	- When multiple technologies are being compared, it may be useful to present the net benefit of each option, and histograms showing the probability with which each one is ranked best (highest net benefit), second, third and so on. - No implementation issues identified.
Categorising uncertainties by their resolvability	Uncertainties presented to committees may differ by how practical it is to resolve them with additional evidence, such as further clinical data or expert advice, in a way that is useful for current decision making.	Yes – minor	- Different types of uncertainty could be categorised, and the different categories can be presented in a simple colour-coded visual framework. - Broadly, this should show uncertainties that were appropriately captured in the analysis, those that might practically be resolved in a reasonable time frame, and those that cannot reasonably be resolved. This could help to focus committee discussions.

Topic	Current methods	Case for change	Comments
			- The working group considered that this could occur after committee meetings, rather than before, as a way of transparently explaining how committee interpreted different uncertainties. - The approach discussed is interesting and highlights a case for change, but further work is needed.
Evidence generation over time	A graphical representation of how the evidence base has developed over time, including when future evidence might become available, could be presented, based on a similar graph used in the aviation industry.	Yes – minor	- Companies should submit a graphical presentation of the evidence for a technology over time. - This should include planned future evidence points (for example, data readouts or further studies), and a description of which uncertainties those data might address. - This may help committees to consider the value of evidence-generating recommendations. - No implementation issues identified. Committees already consider future evidence, but without a visual aid to help.

Table 6 Expert elicitation: findings and proposed cases for change

Topic	Current methods	Case for change	Comments

Eliciting expert opinion	This topic was not formally considered by this group. However, the group would like the incorporation of formal expert elicitation in the Centre for Health Technology Evaluation processes to be considered by the process review.	No	- Formal expert elicitation could be used to agree on plausible parameter ranges and scenarios, and to define which uncertainties are unresolvable. - Current processes do not allow this to occur during an appraisal or evaluation. - Companies can inform their submissions with formal expert elicitation.

Introduction

3. Uncertainty is present in every health technology assessment. There are generally 3 types of uncertainty that are considered during decision making: **choice of data source**, **parameter uncertainty** and **structural uncertainty**. Together, these elements contribute to the overall **decision uncertainty** faced by NICE committees. In general, greater decision uncertainty makes the true clinical and cost effectiveness of an intervention less clear, which may therefore make it less clear which choice is optimal. If NICE makes a suboptimal recommendation, either positive or negative, it imposes opportunity costs on the health and care system.

4. NICE committees are often asked to make decisions about technologies with limited clinical- and cost-effectiveness evidence. This has long been true of diagnostics and medical technologies, and it is becoming increasingly common for pharmacological products due to earlier licensing decisions, the demand for early access and an increased number of treatments targeting smaller patient populations, such as those for rare conditions. It is therefore necessary for CHTE to ensure its methods for handling uncertainty remain appropriate. This is reflected in the **2019 Voluntary Scheme**: 'NICE will clarify its approach to managing uncertainty in the appraisal of a new technology, brief its committees on the types of uncertainty and ensure that committee discussions focus on those areas of uncertainty that have the most significant impact on estimates of cost effectiveness.'

5. Accordingly, this report reviews the current approaches used within CHTE and methods recommended by other international HTA bodies (identified through targeted searching and cross-referencing to the EUnetHTA 'methods for health economic evaluations' report [2015]). The report also considers key academic literature (identified by targeted searching and approaching key academics for key papers) and guidance from the Treasury. The report is also informed by workshops to discuss how uncertainty is presented in other (non-HTA) sectors and value of information analysis.

6. The following issues are considered in this report:

- Probabilistic analysis

- Threshold analysis

- Structural uncertainty

- Extrapolating beyond the available data

- Value of information analysis

- Presenting uncertainty

  The report considers the role of each of these issues in informing decision making. It details areas in which the task and finish group believes a potential case for change to CHTE methods exists. Where appropriate, cases for change and proposed wording for the methods guides are presented for consideration. The report also details any expected practical challenges associated with implementing that change across CHTE.

7. The conclusions from this report apply to NICE's Technology Appraisals (TA), Highly Specialised Technologies (HST), Diagnostics Assessment (DAP) and Medical Technologies Evaluation (MTEP) Programmes.

8. The extent to which this report addresses the issues and questions in the project specification is detailed in the Appendix.

Overlaps with other elements of the methods review

9. It is recognised that several elements of the CHTE methods review interact with each other. This document should be considered in conjunction with other reviews within the methods update. For this topic, the key related reviews are:

   - **Evidence sourcing and synthesis**: this group will consider choice of data source uncertainty associated with identifying and synthesising the available evidence, and methods to elicit expert judgement.

   - **Modifiers**: this group will consider the role of uncertainty as a modifier for altering the decision-making threshold, and how it should interact with other modifiers.

   - **Overall approach to decision making**: this group will consider how uncertainty should inform a committee's final decision.

Probabilistic analysis

10. Decision-making committees within CHTE are typically presented with a 'base-case' estimate of a technology's cost effectiveness, representing the estimate considered to be most plausible by the company, academic group or NICE technical team, given the evidence available. Such base-case estimates are derived from cost-effectiveness models or cost-consequence analyses that need various input parameters (such as health outcomes, utilities and resource use). Often, the mean value of a given input parameter provides the most appropriate point estimate to be used in a deterministic analysis. Parameter uncertainty describes uncertainty in the 'true' value of those inputs; if a parameter is highly uncertain, there may be a wide spread of values that it could plausibly take in practice. While 'extreme' values further from the appropriate point estimate may be less unlikely to occur, they could have a large effect on cost effectiveness.

11. Probabilistic analysis attempts to characterise parameter uncertainty by assigning a plausible distribution to each input, drawing random values from those distributions and calculating the resulting cost and health outputs, then repeating this a large number of times to estimate the mean and variance of cost and health outputs. If a model's probabilistic results are dissimilar to its deterministic results then, assuming it has been implemented correctly, the model is 'non-linear'. Here, at least 1 uncertain parameter causes a bigger change to cost-effectiveness results when it is changed in 1 direction from its mean value (for example, gets higher) than the other direction (for example, gets lower). Here, the probabilistic results will provide a more accurate estimate than deterministic results. Therefore the probabilistic estimate should be used for decision making.

Current Centre for Health Technology Evaluation (CHTE) methods

12. The TA, HST and DAP method guides are clear that parameter uncertainty should be explored using probabilistic analysis, either as sensitivity analyses, or as the preferred analysis if a model is shown to be non-linear. The MTEP guide is less prescriptive, stating that probabilistic analysis may be used if it is appropriate given the complexity of the decision problem. Medical technologies typically have weaker evidence bases (based on the conventional hierarchy of evidence). So, MTEP's approach may reflect the increased difficulty in robustly parameterising inputs.

Probabilistic analysis in decision making

13. The review identified 2 aspects of probabilistic analysis that needed consideration: (a) getting the most accurate estimate of cost effectiveness and (b) how probabilistic results are used in decision making.

14. While probabilistic analysis is included in all CHTE methods guides, and is explicitly needed in all programmes except MTEP, it does not appear to be consistently used to inform decision making, even within individual programmes. It is common for analyses to be primarily deterministic after probabilistic results are shown to be similar to base-case results. In part, this may be because academic groups have limited time to critique company analyses; reviewing the application of probabilistic analysis in close detail, and re-running all results probabilistically, may mean trading off other important aspects of their critiques. Instead, committees often focus on key deterministic scenario analyses in making a final decision. This may be inaccurate; if any scenario is non-linear, then the deterministic result will be less accurate than the probabilistic result, so may lead to incorrect conclusions about cost effectiveness. It will also tend to lack useful information for decision making, such as the probability that the preferred cost-effectiveness estimate is within a given threshold, and the likelihood of that scenario occurring (deterministic scenario analyses simply show an alternative 'state of the world'). The likelihood of an input parameter taking a particular value is implicitly captured by the probability distributions used in a probabilistic analysis (but for alternative scenarios, expert and committee judgement is needed to decide which is the most suitable analysis; see paragraphs 78 to 81).

15. When cost-minimisation analysis is accepted for decision making, such as in a fast-track TA, a probabilistic analysis is not necessary because the decision problem becomes a simple assessment of which option has the lowest known cost.

Developments from policy, academia and other health technology assessment (HTA) bodies

16. The extent to which probabilistic analysis is preferred varies widely across international HTA bodies. At one end of the spectrum, various HTA bodies do not expressly need probabilistic analysis (including **Denmark**, **Portugal**, **Slovenia**). Perhaps most notable among them is the HTA body in **Scotland**, which states that probabilistic analysis is not mandatory because robust evidence to parameterise all model inputs may not be available before a technology has been used in practice. Taking the opposite stance is the HTA body for **Canada**, which gives clear primacy to probabilistic analysis, stating that deterministic analysis is inappropriate because it cannot fully reflect correlation between parameters. Its methods therefore also avoid using the otherwise common-term probabilistic sensitivity analysis, on the grounds that analyses should be probabilistic by default. Most HTA bodies, including NICE, stipulate that some level of probabilistic analysis should be presented,

at least to explore parameter uncertainty through probabilistic sensitivity analysis.

17. Current Treasury guidance on evaluation and producing analysis does not strictly need probabilistic analysis in all cases. The Treasury **Aqua Book (2015)** ('guidance on producing quality analysis for government') states that parameter uncertainty can be quantified by sensitivity analysis or probabilistic modelling. The Treasury **Green Book (2018)** ('central government guidance on appraisal and evaluation') states that probabilistic analysis can be used to understand the impact of uncertainty on estimated outcomes. However, the fact that both books note the potential usefulness of probabilistic analysis indicates it may be expected that analysts do at least consider using it for their decision problem.

Cases for change to CHTE methods

18. Probabilistic analysis is not always needed by HTA bodies, nor held in the relatively high regard that it has in current TA and DAP methods. However, removing the need for NICE committees to consider probabilistic analyses would be a retrograde step in its methods. Relying on deterministic results would mean accepting that an unknown quantity of parameter uncertainty is present for each decision problem. Furthermore, if a model is non-linear, we can be certain that a deterministic analysis will not provide the most accurate estimate of cost effectiveness. Despite this, current decision making by committees often relies heavily on deterministic scenario analyses. A likely consequence is that probabilistic analyses may have had suboptimal scrutiny during the evidence critique part of past appraisals and evaluations. Ultimately, this increases decision uncertainty and the risk of decision-making error. Therefore, it is proposed to encourage more frequent consideration of probabilistic outcomes across CHTE.

19. We recommend that CHTE's methods do not go as far as the Canadian HTA body, where only probabilistic analyses are considered for decision making. There remains a role for deterministic sensitivity analysis. It may be useful to identify the sensitivity of a cost-effectiveness estimate to 1 parameter, or a small number of closely correlated parameters. Doing so may guide committees in focusing their deliberations, further analysis, expert elicitation or evidence generation recommendations. For decision problems with very limited evidence, which may be the case for technologies for very rare conditions, it is recognised that there may be high level of uncertainty about a high number of parameters. While this should be characterised, in these circumstances key scenario analyses may still be relatively more important for decision making. This may also be more common for novel medical technologies; however, we feel the current distinction between MTEP and the other CHTE methods guides is unwarranted. In principle, there is no reason

that a starting position of 'use probabilistic results whenever it may be useful for decision making' should not apply across CHTE (except when a pure cost-minimisation analysis is considered an acceptable basis for decision making).

20. The following text is proposed for the updated CHTE methods guide:

For decision problems that use cost–utility, cost-effectiveness or cost-consequence analyses, the committee's preferred estimate should be derived from a probabilistic analysis when possible. If deterministic model results are used, this should be clearly justified, and the committee should take a view on whether the deterministic or probabilistic estimates are most appropriate. However, in general, uncertainty around individual parameters is not a reason to exclude them from probabilistic analyses; rather, that uncertainty should be captured in the analysis.

In general, scenario analyses should also be probabilistic. When only deterministic base-case or scenario analyses are provided, this should be justified. For example, it may be impractical for assessment groups to get probabilistic results for many plausible scenarios. This may be less influential for decision making if the base-case analysis is shown to be linear, or only moderately non-linear (when 'non-linear' means that there is not a straightforward linear relationship between changes in a model's inputs and outputs).

Probabilistic analysis is not needed for decision problems in which the results of a cost-minimisation analysis are sufficient for decision making. Then, the decision should be based on a comparison of known (certain) costs only. These are not subject to parameter uncertainty.

The committee's judgements on cost effectiveness should consider the spread of results, which can be presented using confidence intervals around costs, quality-adjusted life years (QALYs) and net benefits, and a cost-effectiveness plane. The committee should also consider the probability that a preferred cost-effectiveness estimate is below the relevant threshold value. For most decisions, threshold values of £20,000 per QALY gained and £30,000 per QALY gained should be used. The maximum expected value of the decision uncertainty should also be presented to committees by calculating the Expected Value of Perfect Information (EVPI).

Deterministic sensitivity analyses exploring individual or multiple correlated parameters may be useful for identifying parameters to which the decision is most sensitive. 'Tornado' histograms may be a useful way to present these results. Deterministic threshold analysis might inform decision making when there are influential but highly uncertain parameters. However, if the model is non-linear, deterministic analysis will be less appropriate for decision making.

21. In the TA programme, academic groups have limited time to critique company analyses. If probabilistic analyses are to be prioritised for use in decision making, it will become more important for academic groups to thoroughly review and comment on the application of probabilistic analysis. The reviewing group's preferred analysis should be run probabilistically, and clinically plausible exploratory analyses should also be run probabilistically unless there is satisfactory reason not to do so (for example, practical limitations, or the model has been shown to be linear). There is a risk that this means review groups spend less time on other aspects of their critique.

22. The proposed new paradigm is that probabilistic analysis should be done. If it is not done, for example, for some scenario analyses because of computational constraints, this should be explained including, if possible, the likely effect on cost-effectiveness estimates. If a company thinks that probabilistic results may be less suitable for decision making, it should clearly justify why deterministic results are more appropriate.

23. If a company does not provide probabilistic results, the academic group critiquing a submission, NICE technical staff or committee might disagree with its rationale. It may not be possible or practical for the academic group to run the necessary probabilistic analyses as part of its exploratory analyses. This issue could be discussed during the decision problem, clarification and technical engagement stages of an appraisal or evaluation. If there is still disagreement, it is likely that both sets of analyses (deterministic and probabilistic) will need to be made available for committee to decide which is more appropriate for decision making. Even if probabilistic cost-effectiveness results show a large 'cloud' of parameter uncertainty (for example, if the clinical evidence is largely uncertain), this becomes known information for the committee to consider, and is particularly useful when accompanied by an estimate of the expected cost of the uncertainty (see paragraphs 113 to 132). If probabilistic results are not presented at all, the extent of this uncertainty is unknown. If the committee requests probabilistic results, this could delay its final decision.

24. It is likely that this change would cause a notable shift in how medical technologies are evaluated in MTEP, for which probabilistic analysis is currently not routinely needed. Across CHTE committees, increased used of probabilistic results may increase the use of the 'probability that an incremental cost-effectiveness ratio (ICER) is below a given threshold' as a factor in their decision making, alongside ICER point estimates.

25. The methods working group considered whether the proposed emphasis on probabilistic results will unfairly affect access to technologies for very rare conditions. This is because rare conditions are generally associated with greater limitations in the clinical evidence base and therefore greater

uncertainty. The uncertainty may be viewed as a negative feature of a decision problem. It was noted that the purpose of using probabilistic analysis to get the best available estimate of cost effectiveness does not necessarily affect rare-disease technologies any more than others. Mechanisms that do not rely on data (namely, expert elicitation) exist that can be used to inform a probabilistic analysis and ensure the base-case result is as accurate as possible. Further, the alternative approach of using only deterministic results makes the strong assumption that we are certain about the values of input parameters. It was agreed that in terms of characterising uncertainty – the main remit of this aspect of the methods review – the proposal does not cause an unintended equalities issue. However, it is acknowledged that how uncertainty and rarity are used as a modifier in decision making could potentially create an equality issue.

Showing model convergence

26. For probabilistic results to be meaningful, a model needs to be 'run' many times to get a mean (average) result. The number of individual probabilistic simulations done to get the mean model outputs is usually reported, and tends to be a large, round figure like 1,000 or 5,000. Because of the random element of probabilistic analysis, 2 sets of mean results from running the same model are highly unlikely to be identical (although they will be identical if the same random number sequence is used). This 'random noise' variance, or Monte Carlo error, is not relevant for decision making, and should be eliminated as far as possible. Running a probabilistic model 5,000 times will produce a smaller variance in the results compared with running it 1,000 times; running the model an infinite number of times will mean its results 'converge' on their true mean. This is clearly impractical to do, but without showing that a model has been simulated enough times for its results to converge, the number of probabilistic runs is arbitrary and may mean avoidable Monte Carlo error is contributing to the decision uncertainty.

Developments from policy, academia and other HTA bodies

27. Various HTA bodies, including **Australia**, **Ireland** and **Netherlands**, stipulate that the number of probabilistic model runs should be large enough for mean results to be stable (that is, converged). The HTA body for **Canada** specifies that at least 5,000 runs will typically be needed for this to happen.

28. **Hatswell et al. (2018)** showed that the confidence intervals around net benefit results can be used to compare variation in the results from different numbers of probabilistic model runs. The variance will decrease as the number of probabilistic simulations increases, but when the variance becomes approximately static, the results can be said to have converged. By reviewing the variance, analysts can decide whether more probabilistic model

simulations are needed to achieve convergence. The authors showed that net benefit results have simpler distributional properties than the more common ICERs, making them easier to use to show convergence; however, net benefit confidence intervals can be converted to ICERs if needed.

Cases for change to CHTE methods

29. If model convergence has not been shown, a probabilistic cost-effectiveness estimate may be subject to substantial avoidable Monte Carlo error; if the model was run the same number times again, the results could be very different. At worst, it could change the cost-effectiveness estimate from being under a given threshold to being over it, or the reverse. A decision-making committee needs to be confident that the probabilistic results it assesses would be closely replicated if the analysis was run again. It can then focus on other issues, such as parameter uncertainty and structural uncertainty, in its decision making.

30. The following text is proposed for the updated CHTE methods guide:

When doing a probabilistic analysis, enough model simulations should be used to minimise the effect of Monte Carlo error. Reviewing the variance around probabilistic model outputs (net benefits or incremental cost-effectiveness ratios [ICERs]) as the number of simulations increases can provide analysts with a way of assessing whether the model has been run enough times or more runs are needed.

31. No implementation issues have been identified for this change, which should be done routinely to minimise random noise in cost-effectiveness estimates. The relevant work would be done by the model developer for an appraisal or evaluation (companies or academic groups). In some cases, this could reduce the overall computational burden by showing a relatively low number of probabilistic runs is sufficient for results to become stable. A supplementary file to the Hatswell et al. study provides an Excel tool for using the method described.

Ordered and correlated parameters

32. Most parameters used as inputs to a model will have logical bounds on what values they can take; for example, a unit cost cannot be less than £0. Clearly, the distributions used to parameterise inputs for probabilistic analysis should ensure that negative costs are not used. For related parameters, the logical bounds might be less obvious. For example, it might be clinically implausible that the quality of life of patients with 'severe disease' would, on average, be better than that of patients with 'mild disease'. Therefore, in a model, the utility value for the 'severe' health state should be lower than the value for the 'mild' health state. Such interdependencies can be reflected in the probabilistic analysis. Common, suboptimal approaches to do this include

independently sampling then subsequently restricting parameters (artificially changing their true distributions) or using the same random number generator for each parameter (artificially imposing perfect correlation between them).

Developments from policy, academia and other HTA bodies

33. A requirement that the correlation between parameters is considered in uncertainty analyses is commonly stipulated by HTA bodies that produce comparatively detailed methods guides, including **Australia**, **Canada, France, Germany**, **Ireland** and the **US**. They do not make explicit reference to considering ordered parameters as part of this.

34. **Ren et al. (2018)** proposed a novel approach to sampling ordered parameters for the purpose of probabilistic analysis (the 'difference method'). The technical details are too involved to be referenced in the CHTE methods guide. In short, the related (ordered) parameters are transformed so they are unbounded, then the difference between them is sampled, rather than the individual values themselves. The authors assert that this method can provide a robust approach to modelling ordered parameters, when more naive existing approaches often lack statistical and clinical validity.

Cases for change to CHTE methods

35. Some parameter inputs are characterised by natural or logical orders; using the earlier example, quality of life with 'mild' disease severity might logically be expected to be better than with 'severe' disease. However, because of uncertainty around their mean values, the distributions of the 2 utility values may overlap, which would lead to some probabilistic model runs using a higher utility value for people with worse quality of life. This increases the overall uncertainty around a mean ICER, more so the more the distributions overlap. While the TA methods guide currently states that correlation between parameters should be considered, commonly used methods to impose a logical order on parameter distributions are suboptimal. Therefore, the methods guide could be supplemented with text that notes these methods should not be used, and more sophisticated approaches should be considered.

36. The following text is proposed for the updated CHTE methods guide (blue, bold text indicates new wording):

Evidence about the extent of correlation between individual parameters should be carefully considered and reflected in the probabilistic analysis. **When considering relationships between ordered parameters, consider approaches that neither artificially restrict distributions nor impose an unsupported assumption of**

perfect correlation. Assumptions made about the correlations should be clearly presented.

37. No implementation issues have been identified for this change.

Implausible parameters

38. The task and finish group considered whether the issue of ordered parameters is part of a wider need to ensure all model inputs are plausible, regardless of the type of analysis being conducted. CHTE methods already state that sensitivity, scenario and subgroup analyses should be conducted when 'plausible' alternatives exist, and that committees must be satisfied that 'assumptions used in the reference-case economic modelling are plausible, objective and robust.' However, the group considered that there were still instances in which clinically implausible values were being included in analyses that were presented to committees for decision making. For example, some types of analysis, such as linear regression, may produces implausible values (for example, utilities above 1) when they are evaluated probabilistically.

Cases for change to CHTE methods

39. To prevent the use of implausible analyses, the following paragraph is proposed:

In general, all model parameter values used in base-case, sensitivity, scenario and subgroup analyses should be both clinically plausible and should use methods that are consistent with the data. Results from analyses that do not meet these criteria will not usually be suitable for decision making.

Sometimes it may be difficult to define what is plausible and what is not, for example, in very rare conditions or for innovative medical technologies, when the evidence base may be less robust. In such situations, expert elicitation should be considered to identify a plausible distribution of values.

If threshold analysis is used, the parameter value at which a cost-effectiveness estimate reaches a given threshold may be implausible, and it remains appropriate to present this information.

40. No implementation issues have been identified for these changes. The relevant work would be done by the model developer for an appraisal or evaluation (companies or academic groups).

41. If the proposed change ensures implausible or inappropriate parameter values are not considered in decision making, committee discussions will be more likely focus on the key areas of uncertainty that have an effect on

estimates of cost effectiveness (a commitment by NICE as part of the 2019 Voluntary Scheme).

42. The group also proposed a case for change to prevent implausible scenario analyses being conducted (see paragraph 81).

Link to process review: expert elicitation

43. The group considers that there is potential benefit from addressing the plausibility of parameter values and modelling assumptions early in the NICE process, for example, at the decision problem and technical engagement steps in the TA programme. The company, academic group and NICE team could establish a 'plausible set' of inputs to take forward in the appraisal or evaluation, using formal expert elicitation to inform plausible parameter ranges and scenarios. This would prevent implausible analyses reaching the committee stage and potentially framing the committee's deliberations. The group agreed that companies can currently use thorough, formal expert elicitation to inform their evidence submissions, and increased use of this would help inform key uncertainties early on. However, some uncertainties may remain, for example, if they only become evident after the academic group critiques the evidence.

44. The task and finish group recognised that for most CHTE programmes, incorporating formal expert elicitation during an appraisal or evaluation would need a significant change from current processes. In the TA and HST programmes, the results of expert elicitation are sometimes included in company submissions, but there is limited scope for formal elicitation after a submission is received. It would therefore like this to be considered in the process review.

45. Late in the development of this report, a task and finish group member alerted the NICE team to a recent publication defining best practices for engagement between multiple stakeholders in the context of rare diseases (**Annemans and Makady, 2020**). The authors proposed a collaborative process to identify potential sources of uncertainty, agree which are resolvable (and which are not), and inform further evidence generation, from early in a technology's development through to health technology assessment and beyond. The NICE team considered that some of this engagement is already captured within CHTE processes (for example, NICE can provide early scientific advice, and offers engagement during decision problem meetings before a company submitting evidence). However, the NICE team felt that the inclusion of formal expert elicitation in the process sits within the paper's topic of improving stakeholder engagement processes in general, and therefore proposes that Annemans (2020) is considered in the process review.

Probabilistic one-way sensitivity analysis

46. One-way, or univariate, sensitivity analysis involves varying model input parameters individually to isolate the consequences of each one on the results. This can identify which inputs cost and health outcomes are the most sensitive to. Its results are often presented as 'tornado' diagrams; vertical histograms showing the upper and lower cost-effectiveness estimates achieved by varying each parameter to given upper and lower bounds. A known limitation of one-way sensitivity analysis is that it may be inappropriate to change the value of one parameter in isolation if it is correlated with others. For example, in conditions in which overall survival and disease-free survival are closely related, it may be implausible to increase one of these parameters in a model without also increasing the other.

Developments from policy, academia and other HTA bodies

47. One-way sensitivity analysis is commonly needed or recommended by other international HTA bodies; it is included in **15 of 27** guidance documents reviewed as part of this methods update. A notable outlier is the HTA body for **Canada,** whose methods guide explicitly recommends against the use of deterministic analyses.

48. **McCabe et al. (2020)** argue that probabilistic one-way sensitivity analysis can overcome the known limitations of deterministic one-way analyses, namely: the minimum and maximum values tested may be arbitrary; it does not inform decision makers about the likelihood of those maximum and minimum values occurring; and it does not typically capture the aforementioned correlation between parameters. The authors present a method to estimate the net benefit associated with input parameters taking specific values – like deterministic one-way sensitivity analysis – while incorporating the likelihood of each value occurring. It therefore provides the likelihood of a parameter taking an observed value that causes positive or negative net benefits, which can be presented using a simple line graph.

Cases for change to CHTE methods

49. The TA methods guides currently recommends that univariate sensitivity analysis is a useful tool for decision making. It recognises that deterministic analyses 'become increasingly unhelpful in representing the combined effects of multiple sources of uncertainty', and that correlation between parameters should be considered. It may be appropriate for highly correlated parameters to be removed from one-way sensitivity analysis and instead captured jointly in multiway sensitivity analysis, or reasonable for probabilistic one-way analyses to be explored. The authors of the McCabe et al. (2020) study provide a supplement with the necessary code to do so.

50. As noted earlier, we feel that there remains a role for deterministic sensitivity analysis in CHTE decision making. It may be useful to quickly identify the sensitivity of a cost-effectiveness estimate to one parameter, or a small number of closely correlated parameters. Doing so may guide committees in focusing their deliberations, further analysis, expert elicitation, or evidence generation recommendations. The current TA methods guide wording around univariate sensitivity analysis should be expanded to note that probabilistic univariate analysis may be useful to explore.

51. The following text is proposed for the updated CHTE methods guide:

Probabilistic univariate sensitivity analysis may be explored to incorporate the likelihood of a parameter taking upper and lower bound values, rather than just presenting the effect of it taking those values.

52. Because this change does not propose that probabilistic univariate sensitivity analysis is needed in all circumstances, no implementation issues have been identified.

Threshold analysis

53. One-way sensitivity analysis is a form of analysis that explores the sensitivity of results, such as cost-effectiveness estimates, to changes in individual input parameters. Threshold analysis is a subtype of one-way sensitivity analysis that identifies the 'switching value' value for an input parameter which would change the decision of the evaluation. For example, this may be exploring a range of health state utility values to determine the value at which the results go from cost ineffective to cost effective at a given willingness to pay threshold. Threshold analysis can be done deterministically on any input parameter. While this is typically straightforward to do in cohort models, it is more challenging in models that simulate individual patient journeys.

Current Centre for Health Technology Evaluation (CHTE) methods

54. The Technology Appraisals (TA) and Diagnostics Assessment Programme (DAP) methods guides both stipulate the use of sensitivity analysis (in various forms) to explore uncertainty. However, neither guide makes explicit reference to threshold analysis. The Medical Technologies Evaluation Programme (MTEP) methods guide (Section 7.3.2) stipulates that uncertainty analysis techniques should be used, and lists threshold analysis as an example of such a technique.

Identifying switching values

Developments from policy, academia and other health technology assessment (HTA) bodies

55. Of the 27 international HTA methods guides reviewed, 7 mentioned threshold analysis in their methods guides (**Germany, Ireland, Portugal, Scotland, Spain, Sweden, US**). This indicates that several other HTA bodies consider the identification of parameter values at which the cost-effectiveness conclusion changes to be informative for decision making. In contrast, the HTA body for **Canada** explicitly recommends against the use of deterministic analyses.

56. In most of the identified methods guides, threshold analysis is only mentioned briefly and is generally listed as a form of sensitivity analysis that can be used. The HTA bodies in **Ireland** and **Portugal** outline that threshold analysis should be used when there are doubts about the accuracy of the data used, and the latter advises that the resulting switching values should be discussed in light of the available economic and clinical evidence.

57. The Treasury's **Green Book** specifies that '[at] a minimum sensitivity analysis and the identification of switching values should be carried out. These results

must form part of the presentation of results.' It also provides a worked example of threshold analysis. It explains that the reason for identifying switching values is to assess the extent to which benefits can fall short of expectations, or costs can exceed expectations, for a proposal (technology) to remain 'value for money'. Once the switching value is known, decision makers can consider the likelihood of that value occurring in reality.

Cases for change to CHTE methods

58. Threshold analysis is already mentioned in the MTEP methods guide and, while it is not explicitly referred to in the TA or DAP guides, threshold analysis has been used to inform committee decision making in these programmes. We propose that threshold analysis is explicitly included in the methods guide for all programmes as an option for exploring highly uncertain parameters. We propose to add wording that notes circumstances in which threshold analysis may be particularly useful for decision making (such as showing the cost-effectiveness estimate is robust to even very extreme parameter values).

59. We note that there are some circumstances in which threshold analysis will not be appropriate, and propose to qualify that:

- Threshold analysis should not be used as a justification for restricting the population of interest to a cost-effective subgroup.

- Threshold analysis may not be appropriate in some circumstances, for example, if used to explore uncertainty around parameters that are highly correlated with other influential parameters in the model.

- Threshold analysis may be impractical to do in models that simulate individual patient journeys.

60. The following text is proposed for the updated CHTE methods guide:

Threshold analysis can be used as an option to explore highly uncertain parameters when identifying a parameter 'switching value' may be informative to decision makers. A switching value is the value an input variable would need to take for a proposed intervention to switch from being cost ineffective to cost effective for a given threshold (for example, £20,000 and £30,000 per quality-adjusted life year gained). The threshold analysis should indicate how far the switching value is from the current best estimate of a parameter value.

Threshold analysis is not be suitable for exploring uncertainty around parameters that are highly correlated with other influential parameters. Threshold analysis should also not be used to justify restricting the population of interest to a cost-effective subgroup.

61. No implementation issues have been identified for these changes.

62. Threshold analysis may become frequently used to show that cost-effectiveness estimates are robust to uncertain parameters, by showing that those parameters would need to take very extreme values to change the conclusion about cost effectiveness. This would help to ensure that committee discussions focus on other areas of uncertainty that have a more important effect on estimates of cost effectiveness (a commitment by NICE as part of the 2019 Voluntary Scheme).

Structural uncertainty

63. NICE methods guides highlight the importance of investigating the effect of structural uncertainty on cost-effectiveness estimates. Examples of structural uncertainty may include how different states of health are categorised and how different pathways of care are represented in models. However, it often gets relatively little attention, especially compared with parameter uncertainty, despite potentially having a major effect on decision making. The proposed amendments to the unified Centre for Health Technology Evaluation (CHTE) methods guide aim to ensure that decisions made by analysts about model structure are clearly set out, to help committees understand what structural assumptions have been made (and why) and ensure that the effect of these assumptions is appropriately investigated. A further proposed change reflects developing methodology in how structural uncertainties can be incorporated in economic models, rather than being assessed independently using multiple scenario analyses.

Current Centre for Health Technology Evaluation (CHTE) methods

64. Both the Technology Appraisals (TA) and Diagnostics Assessment Programme (DAP) method guides stipulate that structural assumptions should be clearly documented, and the evidence and rationale to support them provided. They state that the effect of structural uncertainty on estimates of cost effectiveness should be explored by separate analyses of a representative range of plausible scenarios. The DAP methods guide states that each alternative analysis should present separate results.

65. The Medical Technologies Evaluation Programme (MTEP) methods guide does not specifically advise on structural uncertainty, but states that uncertainty analysis techniques (relating to chance, evidential and model uncertainty) should be done. The level of complexity should be appropriate for the specific technology and its comparator healthcare pathway.

Model conceptualisation

66. Economic models are not always accompanied by a transparent and detailed account of how the model was developed, including model conceptualisation. This includes decisions about which health states to include, whether an individual patient simulation should be done and, importantly, the evidence and support for these structural assumptions. Greater transparency on how a model's structure was developed would allow committees to have greater confidence that all structural assumptions that have been made are identified, and, if possible, the effect of these assumptions on cost-effectiveness estimates has been explored.

Developments from policy, academia and other health technology assessment (HTA) bodies

67. **Afzali et al. (2018)** highlighted that there is a greater focus on model conceptualisation in recent **Australian** and **Canadian** HTA guidance. The authors also concluded that, in submissions to national funding bodies, the conceptual model development process and all other underlying structural choices and assumptions should be explicitly reported and justified. They suggest written documentation should capture the proposed model structure, the process to identify evidence to inform structural aspects, factors influencing the choices and assumptions, expert consultations and graphical presentations such as influence diagrams.

68. In an article adapted from the NICE Decision Support Unit (DSU) Technical Support Document (TSD) 13, **Tappenden (2014)** provided guidance about formal processes through which an appropriate model structure should be determined. The author noted that, in practice, the reporting of model structures tends to be very limited and, if present, usually focuses only on the final model that has been implemented. This leads to uncertainty about whether the selected model structure is credible; which evidence has been used to inform its structure; why certain abstractions, simplifications and omissions have been made; why certain parameters were selected for inclusion (and why others have been excluded); and why the included parameters have been defined in a particular way. This work highlighted the importance of clinical opinion in deciding model structure, and that decisions made should be clearly documented and reported.

69. A report by the International Society for Pharmacoeconomics and Outcomes Research Modelling Good Research Practices Task Force (**Roberts et al. 2012**) on model conceptualisation provided several recommendations for good practice:

- The modelling team should consult widely with subject experts and stakeholders to ensure that the model represents disease processes appropriately and adequately addresses the decision problem.
- The problem conceptualisation should be used to identify key uncertainties in model structure in which sensitivity analyses could inform their effect.
- An explicit process (expert consultations, influence diagrams, concept mapping, or similar method) should be used to convert the problem conceptualisation into an appropriate model structure, ensuring it reflects current disease knowledge and the process modelled.

Cases for change to CHTE methods

70. There is a general expectation in CHTE methods guides that structural assumptions should be clearly documented and the evidence and rationale to support them provided. Supplementing this with a statement that the model conceptualisation process should be explicitly described should ensure greater focus on this in model reports, and provide committees with greater awareness about the structural assumption made to model complex diseases and care pathways.

71. The following text is proposed for the updated CHTE methods guide (blue, bold text indicates new wording):

Structural assumptions should be clearly documented, and the evidence and rationale to support them provided. **The conceptual model development process used to inform the choice of model structure should be transparent and justified. This should include details of expert involvement in this process (for example, number of experts, details of their involvement, how they were chosen). It is not sufficient to state that the chosen model structure has previously been used in published model reports or accepted in submissions to NICE. The chosen type of model (for example, Markov cohort model, individual patient simulation) and model structure should be justified for each new decision problem.**

72. If model development decisions are laid out in a transparent way, including the rationale and expert involvement, committees are likely to be more comfortable in accepting the model is structurally acceptable for decision making. Committee discussions would then be more likely to focus on other areas of uncertainty that have a more important effect on estimates of cost effectiveness (a commitment by NICE as part of the 2019 Voluntary Scheme).

73. No implementation issues have been identified for this change. Modellers should already have a logical process to decide on an appropriate model structure, including consulting clinical experts, so no additional work will be needed. Guidance to support this process is available in TSD 13 from the DSU.

Investigating structural uncertainty

74. CHTE assessments currently use scenario analyses to investigate structural uncertainty. This is consistent with TA and DAP methods guides, which state that structural uncertainty should be investigated using separate analyses of a representative range of plausible scenarios.

Developments from policy, academia and other HTA bodies

75. **Sculpher and Palmer (2020)** highlighted that there have been recent developments in methods to bring together structural and parameter uncertainties using probabilistic analysis. The authors noted that these methods have not been given any role in NICE methods guidelines, which have instead indicated a preference for the use of scenario analysis.

76. **Afzali et al. (2018)** noted that the HTA body in **Australia** recommends parameterising structural assumptions when there is sufficient clinical evidence or expert opinion to do so (and using scenario analyses if not). The authors concluded that, if alternative structural assumptions are indicative of substantially different model predictions, a formal approach to characterise that uncertainty should be used. In the absence of data to inform an appropriate probability distribution (that is, no prior information and no sufficiently reliable expert beliefs), or when a particular structural aspect cannot meaningfully be parameterised, revert to scenario analyses. HTA guidance in **France** also recommends that, if scenario analyses suggest different decisions may be appropriate, a method of parameterising structural assumptions – model averaging – should be explored.

77. **Jackson et al. (2011)** commented that structural uncertainties are often only explored informally through scenario analysis. They commented that presenting multiple models (for example, scenario analyses) can lead to implicit averaging, without consideration of relative plausibility of each model, or selecting just 1 model to use. The authors showed 2 approaches that can be used to formally account for structural uncertainties, based on whether data to inform parameter likelihood is available or not (in which case expert elicitation can be used). Simple scenario analyses were recommended for use if no information on those likelihoods is available.

Cases for change to CHTE methods

78. There are likely to be advantages to formally incorporating structural uncertainty in probabilistic models, when possible, over multiple scenario analyses. This would allow a single cost-effectiveness estimate to represent structural uncertainties, rather than needing a potentially large number of different estimates. Doing so may also prevent exploratory scenario analyses that are highly unlikely to occur being given undue weight in decision making. It may therefore improve decision making and make NICE processes more methodologically credible. Incorporating structural uncertainty as parameters in a model would allow this to be included in value of information analysis (that is, to inform if research to resolve the structural uncertainty is worthwhile).

79. It may not be possible or feasible to parameterise structural uncertainty in all instances, and simple scenario analyses may need to be reverted to despite considering more formal methods. Therefore, any proposed change to the methods guide should not stipulate that parameterisation of structural uncertainty must be done, nor that a particular method should be used. Rather, it should highlight that it may be explored or used, as a relaxation of the current requirement for scenario analysis.

80. There is a concern that parameterisation of structural uncertainty could be inappropriately used. For example, presentation of several implausible scenarios could intentionally be included to make a favourable scenario look more likely. Another example would be using goodness of fit to data (AIC/BIC) to assign probabilities to different models for extrapolating survival, which would disregard the clinical plausibility of the long-term extrapolations.

81. The following text is proposed for the updated CHTE methods guide (blue, bold text indicates new wording):

The effect of structural uncertainty on estimates of cost effectiveness should be explored by separate analyses of a representative range of scenarios **that are plausible and consistent with the evidence. Analyses based on demonstrably implausible scenarios are only useful if they are used to show that cost-effectiveness estimates are robust to a source of uncertainty. For example, if the resource use associated with a procedure is uncertain, a useful exploratory analysis might show that the implausible assumptions of no resource use and very large amounts of resources do not materially affect the cost-effectiveness conclusion. The purpose of such analyses should be clearly presented. This will allow a committee to focus on other, key uncertainties in its decision making.**

It may be possible to incorporate structural uncertainty within a probabilistic model (for example, by model averaging or assigning a probability distribution to alternative structural assumptions). If structural uncertainty is parameterised, careful consideration should be given to the alternative assumptions and any probabilities used to 'weight' them. This should be transparently documented, including details of any expert advice.

82. The proposed change means the purpose of exploratory scenario analyses should be made clearer, with demonstrably implausible scenarios only presented if they are useful for decision making by showing robustness. This should help to ensure that committee discussions focus on other areas of uncertainty that have a more important effect on estimates of cost effectiveness (a commitment by NICE as part of the 2019 Voluntary Scheme).

83. Greater adoption of parameterising uncertainty would potentially increase the workload for companies and assessment groups producing models. For many structural uncertainties, robust data to inform the probability that each alternative is appropriate may not be available, therefore expert elicitation may be needed. This would be resource intensive. Also, there may be uncertainty in which experts to use, how or whether to assess their expertise (to inform the resulting probability weights) and the methodology used to synthesise their views.

Extrapolating beyond the data

84. The NICE reference case stipulates that the time horizon for a cost-effectiveness analysis should be long enough to capture all important differences in cost and health outcomes between technologies under evaluation. This frequently poses a challenge because the relevant time horizon is usually substantially longer than the duration of follow up in the relevant clinical-effectiveness evidence. This is particularly common for pharmacological technologies. Technology appraisal (TA) committees have become very familiar with discussing how best to 'extrapolate' beyond the available data to generate a robust cost-effectiveness estimate across an appropriate time horizon. Time-to-event data, such as survival, tends to be the focus of extrapolation modelling, but it is not limited to time-to-event parameters as such. Predicting how data might look in the future is inherently uncertain, and often the principal source of uncertainty in a health technology assessment (HTA).

Current Centre for Health Technology Evaluation (CHTE) methods

85. Both the TA and Diagnostics Assessment Programme (DAP) method guides stipulate that any extrapolation of model inputs beyond the observed data should be reported transparently and exhibit clinical validity. They state that alternative extrapolations should be considered, citing a measure of relative treatment effect as an example. If a treatment effect is derived from evidence with a shorter duration than the appropriate time horizon, potential extrapolation scenarios include: assuming no continued benefit after the duration of follow up; assuming the observed treatment effect is maintained beyond the duration of follow up; and assuming the treatment effect diminishes in the long term. Extrapolation is not mentioned in the current Medical Technologies Evaluation Programme (MTEP) methods guide.

Extrapolating treatment benefits

86. Comments from the task and finish group about current CHTE methods for extrapolating a treatment effect duration beyond the available data suggest that the examples used in the text might lack plausibility. In particular, the group questioned whether assuming there is no benefit after the follow-up duration is a plausible assumption, particularly as it is unlikely that people would cease treatment if it continued to be effective. The group considered that, if this scenario is not plausible, then it is questionable whether it should be presented at all. The existing text is quite blunt and may benefit from added context.

Cases for change to CHTE methods

87. Considering these concerns, the following text is proposed for the updated CHTE methods guide (blue, bold text indicates new wording):

For duration of treatment effects, scenarios in the extrapolated phase might include:

- **treatment effectiveness ceases or diminishes gradually over time**

- **treatment effectiveness is sustained for people who remain on treatment**

- **some lasting treatment effectiveness is sustained for people who stop treatment, when it is clinically plausible for lasting benefit to remain.**

88. No implementation issues have been identified for this change.

Flexible survival models

89. Historically, 'standard' parametric approaches have been used to fit smooth mathematical functions (curves) to observed time-to-event data, such as survival data. Such curves are not restricted to the duration of the observed data so can be used to model what longer-term survival might look like. Several 'standard' functions might be fitted to the data (for example, exponential, log-normal, Weibull), with the most appropriate curve determined based on internal and clinical validity. The advent of 'flexible' methods for fitting and extrapolating survival data (for example, spline, mixture models), which are not limited to the standard forms and can incorporate external data, means TA committees have increasingly been presented with novel approaches. It has been hypothesised that flexible approaches may be better equipped to predict outcomes with new technologies, such as immunotherapy oncology treatments, which may show delayed effects or even a cure for some patients.

Developments from policy, academia and other HTA bodies

90. In general, methods guides published by HTA bodies in other jurisdictions provide limited technical guidance on extrapolating beyond available data. Several use similar methods to the current TA methods guide (including Belgium, Canada, France, Ireland, Netherlands and Scotland). However, guidance from HTA bodies in **Australia** and the **US** explicitly state that more flexible approaches, such as piecewise and cure models, should be explored.

91. Using data from a non-small cell lung cancer trial, **Ouwens et al. (2019)** found that flexible (cure, mixture and landmark) approaches fitted to an interim data cut predicted long-term survival better than standard approaches, though differently to each other. **Bullement et al. (2019)** and

Othus et al. (2017) found a similar result using data from immunotherapy trials in melanoma, with flexible models that incorporated external data (a 'cured' proportion or background survival data) doing best and producing very different cost and quality-adjusted life year outcomes.

92. These studies, and a further study by **Guyot et al. (2017)**, showed that some flexible methods appear to be able to incorporate external data without compromising statistical fit to the observed data. **Jackson et al. (2017)** reviewed methods used to incorporate external data into survival modelling, concluding that doing so is practical to do and potentially more precise.

93. **Bell Gorrod et al. (2019)** reviewed TA oncology guidance since the publication of the Decision Support Unit (DSU) Technical Support Document (TSD) 14 on survival analysis in 2011 and found that only 7% followed all DSU recommendations for extrapolating survival. The validity of extrapolations was a significant factor in the evidence critique in 71% of technology appraisals. The DSU is due to publish a new TSD advising on flexible survival analysis methods.

Cases for change to CHTE methods

94. The TA programme is increasingly presented with flexible approaches to extrapolating beyond available survival data; however, they are not mentioned in current CHTE methods guidance. While the guides state that alternative extrapolations should be explored, we propose that they are made more prescriptive by stating that flexible methods should be considered as part of this. The currently available approaches could be listed (standard parametric, flexible parametric), but this would be explicitly non-exhaustive, to prevent excluding future novel approaches. Extrapolation of time-to-event data is unlikely to be needed in MTEP assessments but, if it is, the same methods would be applicable.

95. The following text is proposed for the updated CHTE methods guide:

When extrapolating time-to-event data, various standard (for example, parametric) and more flexible (for example, spline-based, cure) approaches are available. Their appropriateness and the validity of their extrapolations should routinely be considered.

96. No implementation issues have been identified for this change. The relevant work would be done by the model developer for an appraisal or evaluation (companies or academic groups), and this is increasingly happening already. For decision making, NICE committees are already becoming accustomed to assessing these more complex extrapolation methods. The forthcoming TSD will provide guidance for assessing which methods are likely to be the most appropriate to use.

Selecting the most appropriate extrapolation

97. When curves are fitted to time-to-event data to allow models to extrapolate beyond those data, the internal and external validity of alternative curves are typically compared, to inform which is the most appropriate.

Developments from policy, academia and other HTA bodies

98. Most HTA bodies whose methods discuss extrapolation need transparent justification of the chosen extrapolation (including **Australia**, **Canada**, **Ireland** and **The Netherlands**). Important parts of this decision making include comparing how well the estimated curves match the data visually, comparing statistical goodness of fit values, and seeking advice from clinical experts. However, there may be other factors that can support selecting the most plausible curve.

99. **Kearns et al. (2020)** simulated survival data with different underlying hazards and compared 7 'standard' parametric curves fitted to them, finding that the variances of the curves behaved differently over time. The confidence intervals for some curves decreased in the extrapolated period (that is, after the simulated trial data follow-up period), which implies increased certainty as the curve moves further and further away from the actual data. This may be implausible and may underestimate the true extent of uncertainty.

100. Secondarily, **Kearns et al (2020)** note that it is easier to interpret the long-term behaviour of alternative extrapolations by comparing their underlying hazard plots over time, rather than visually inspecting their survival plots. Differences in survival curves may be hard to distinguish, particularly when the proportion of people remaining alive is low, whereas seeing the mortality hazards faced by people who are still alive over time can indicate differences and implausible assumptions. In their review of survival analyses in oncology TAs, **Bell Gorrod et al. (2019)** noted that only 38% of technology appraisals explicitly assessed overall survival hazard functions.

Cases for change to CHTE methods

101. Companies, academic groups and committees already spend significant time and attention on selecting appropriate time-to-event curves that provide plausible extrapolations beyond. While clinical plausibility of extrapolation remains a paramount consideration, in the event of 2 or more curves being equally plausible, the variance of the curves' extrapolation should be considered in this decision making. Choosing a curve that has constant or decreasing confidence intervals over time is more likely to underestimate the true uncertainty in the extrapolation, relative to the more conservative approach of selecting a curve that has an increasing confidence interval.

102.   Hazard functions over time should be explicitly compared when determining which curve provides the most plausible extrapolation, in all cases. When the most appropriate curve is likely to be an important issue for decision making, hazard functions should routinely be presented to decision-making committees.

103.   The following text is proposed for the updated CHTE methods guide:

When comparing alternative models for extrapolating time-to-event data, the clinical plausibility of their underlying hazard functions should routinely be assessed. Uncertainty in the extrapolated portion of hazard functions should also be explored. Functions that display stable or decreasing variance over time are likely to underestimate the uncertainty in the extrapolation.

104.   As part of the 2019 Voluntary Scheme, NICE has committed to ensuring committee discussions focus on the most important areas of uncertainty. When time-to-event evidence is used in an analysis, the choice of extrapolation is commonly a key factor affecting cost effectiveness. If the proposed changes improve the selection of extrapolation analyses and how they are presented to committees, it should improve decision making for this often-important issue.

105.   No implementation issues have been identified for these changes. While some NICE committee members may be more accustomed to comparing alternative survival functions, hazard functions can be explained by academic groups or NICE technical teams, are no more difficult to interpret.

106.   The requirement for companies to present hazard plots, in addition to survival plots, when providing time-to-event evidence, should be included in company submission templates.

Uncertainty introduced by adjusting for treatment switching

107.   Treatment switching in clinical trials occurs when people in 1 arm 'crossover' to have the intervention in a different arm. This increases uncertainty in resulting time-to-event data, but several statistical methods are now available and routinely used to adjust survival estimates to account for crossover. While treatment switching is not typically discussed in HTA bodies' methods guides, CHTE's TA guide states that when an adjustment method is warranted, the choice of method should be explored and justified.

Developments from policy, academia and other HTA bodies

108.   In the context of using surrogate data, the Treasury's **Aqua Book (2015)** and **Australia's** HTA guidance state that underlying statistical

approaches (like crossover adjustment) can introduce their own 'translational' uncertainty, and that this should be quantified. The TA methods guide does not mention what to do about additional uncertainty that may be introduced by adjusting for trial crossover.

109. **Bennett et al. (2017)** assert that 1 method, the Rank-Preserving Structure Failure Time (RPSFT) model, itself introduces additional uncertainty because it adjusts the original data to account for crossover. Using a simulated oncology trial dataset, they compared the survival estimates from 2 approaches that explicitly account for the fact an RPSFT adjustment has been done with the conventional way RPSFT adjustment is typically used. Both alternative approaches produced wider confidence intervals around survival estimates and were more conservative about the probability of an intervention being clinically superior to placebo.

Cases for change to CHTE methods

110. The TA methods guide makes the following statement about uncertainty when using surrogate outcomes: "In all cases, the uncertainty associated with the relationship between the end point and health-related quality of life or survival should be explored and quantified." Methods for adjusting uncertainty are now so commonly used that a similar statement may be reasonable, advising that additional uncertainty caused by using an adjustment method should be captured. Crossover is unlikely to be present in DAP or MTEP evaluations, but this statement would be appropriate in those settings if needed.

111. The following text is proposed for the updated CHTE methods guide:

When appropriate, the uncertainty associated with the use of a method to adjust for trial crossover should be explored and quantified.

112. No implementation issues have been identified for this change. The relevant work would be done by the model developer for an appraisal or evaluation (companies or academic groups), and would form part of the sensitivity and scenario analyses that are routinely done.

Value of information analysis

113.   Decision-making committees in the Centre for Health Technology Evaluation (CHTE) are increasingly being asked to assess technologies with immature or emerging evidence bases, which can contribute to substantial decision uncertainty. Value of information analysis has been identified as a potential method that could be used to help committees decide whether recommendations for further evidence generation may reduce decision uncertainty when assessing technologies. It involves quantifying the per-patient value of having better information about some or all parameters. While probabilistic analysis informs the likelihood of a choosing a suboptimal outcome, value of information analysis quantifies the expected cost of doing so (or the value of eradicating some or all uncertainty).

114.   Forms of value of information analysis include the Expected Value of Perfect Information (EVPI; estimating the value of perfect information for all parameters), the Expected Value of Partially Perfect Information (EVPPI; estimating the value of perfect information for one or some parameters) and the Expected Value of Sample Information (EVSI; estimating the value of information from a given sample). No research design can be expected to deliver perfect information. Instead, research is likely to deliver imperfect sample evidence on a group of parameters. For this reason, EVSI will always be lower than EVPI. EVPPI will also be lower than EVPI because it is the value of perfect information on only a subset of parameters.

115.   EVPI per patient can be calculated as a simple extension to a probabilistic analysis. While EVPPI per patient is more complex to calculate than EVPI, it can still be implemented fairly easily in many situations, using tools such as the Sheffield Accelerated Value of Information tool. Estimating EVSI per patient is more challenging because it needs information about the research design intended to generate further evidence, although a good practice guide has recently been published (**Kunst et al., 2020**). Evidence generation will have value for decisions about treatment for current and future patients. Because of this, estimates of EVPI, EVPPI and EVSI should all be scaled up to the size of the beneficiary population. This needs epidemiology data on the size of prevalent and incident populations likely to have treatment instead of the alternative option. It also needs an estimate of the 'decision relevance horizon'; the time horizon over which information generated by research is expected to be useful for decisions about treatment choices. The decision relevance horizon is generally subject to substantial uncertainty and can have a large effect on resulting value of information estimates (**Kim et al., 2020**).

Current Centre for Health Technology Evaluation (CHTE) methods

116.   The Technology Appraisals (TA) and Highly Specialist Technologies (HST) Programmes have the option to make recommendations for approval with research via managed access agreements, involving evidence generation and commercial arrangements, for cancer drugs and highly specialised technologies. Cancer drugs may be considered for inclusion in the Cancer Drugs Fund if they show plausible potential for cost effectiveness alongside other criteria. The TA and HST programmes also have the option to make recommendations for technologies to be used only in research, but in practice this option is rarely issued (with only 2 such recommendations made in the past 5 years). The TA methods guide does not currently mention value of information analysis as a potential tool for decision makers to use when considering further evidence generation.

117.   The Medical Technologies Evaluation Programme (MTEP) has the option to make research recommendations within medical technologies guidance. Analogous to the 'only in research' option in TA, MTEP guidance can recommend that a 'case for adoption is not supported but has potential, with recommendation for use in a research context'. Decision makers can also conclude that a technology is 'partially supported in specific circumstances and recommendation for development of further evidence'. However, this recommendation option has not yet been used. When considering research recommendations, the MTEP methods guide (section 8.3) says that when deciding whether to recommend future evidence generation and data collection, the committee can consider:

   a)   'the most important evidence gaps relating to the uncertainty about the technology, and the value of information that could be derived from generating evidence to address them

   b)   information about ongoing or planned research on the technology

   c)   ethical or practical aspects of conducting further research

   d)   the likely costs and benefits of the research (to ensure that a research recommendation does not become a barrier to innovation).'

118.   The Diagnostics Assessment Programme (DAP) manual includes the options to make recommendations for technologies to be used only in research or, in some circumstances, approval combined with a recommendation for further research. It stipulates that these are options are available 'if there is not sufficient evidence to determine the cost effectiveness of the technology'. The DAP programme frequently issues 'only in research' recommendations and issues 'approval with research'

recommendations less commonly. When considering research recommendations, the DAP programme manual (section 16.4) says that when deciding whether to recommend future evidence generation and data collection, the committee considers:

- the most important evidence gaps relating to the uncertainty about the technology, and the value of information that could be derived from generating evidence to address them

- information about ongoing or planned research on the technology

- ethical or practical aspects of conducting further research.

In practice, when considering research recommendations DAP decision makers consider the most important evidence gaps relating to the uncertainty. Approaches used to identify key parameters include threshold analysis, one-way sensitivity analysis and scenario analyses. EVPI has been used once. DAP committees also consider research recommendations about the care pathway after a diagnostic test if uncertainties about the pathway affect the value of testing. Research into diagnostic tests often involves substantial sunk costs as expensive testing machinery may need to be purchased. Adopting diagnostic technologies during a research period can substantially affect care pathways and so may incur costs (such as time or staff fatigue with change).

119.   Literature exploring value of information analysis highlights that the estimates of EVPI, EVPPI and/or EVSI should be considered alongside the cost of generating further evidence. In current CHTE methods, both the DAP and MTEP guides state that the practicalities of conducting research should be considered, and the MTEP guide explicitly mentions the cost of research. However, they do not state that costs should be formally included in value of information calculations; rather, costs appear to be included among general, practical considerations, perhaps because NICE does not fund evidence generation. Instead, value of information analysis is used to explore whether further evidence generation could reduce uncertainty and improve the likelihood of making an optimal decision.

Using value of information in decision making

Developments from policy, academia and other health technology assessment (HTA) bodies

120.   In general, methods guides published by HTA bodies do not state that the value of information analysis can or should be used to inform decision making. Two guides were identified that do explicitly need it in their

reference-case analyses. In **Ireland,** EVPI should be presented to show the value of eradicating all uncertainty. In **Canada,** EVPPI should be presented for parameters identified as being key to the decision when a decision problem includes the prospect of further research.

121. The use of value of information analysis was previously considered in NICE's Strategic Technology Appraisal Review in 2015/16. During this project, NICE commissioned a Decision Support Unit (DSU) report outlining a framework for analysing uncertainty and risk in the context of managed entry agreements (**Grimm et al. 2016**). In addition to considerations about pricing, the report explored the use of concepts such as the incremental expected net benefit gain from adopting a technology expected to be cost effective compared with the next best option (defined as the 'payer optimality gain'). The report also outlined the reciprocal concept of the incremental expected net benefit loss from adopting a technology not expected to be cost effective compared with a cost-effective option ('payer strategy burden'). The expected net benefit gain or loss from adopting a technology that is or is not expected to be cost effective can be compared with the value of information to guide decision making about whether further evidence generation would add value. The report also outlines the requirement that estimates of EVPI, EVPPI or EVSI be scaled up to the size of the beneficiary population over the decision relevance horizon. The Strategic Technology Appraisal Review concluded that value of information analysis should be considered in the current methods review.

122. **Claxton et al. (2016)** outline several other factors that need to be considered when using value of information analysis to inform evidence generation. These include the presence of irrecoverable costs, the likelihood of research being conducted, the length of time for results to report, the opportunity costs of resources used in research and how much of the uncertainty is likely to be resolved through evidence generation. The paper also highlights that there are some sources of uncertainty that may be resolved over time without further evidence generation, such as future price changes.

123. A report by the International Society for Pharmacoeconomics and Outcomes Research Value of Information Analysis Emerging Good Practices Task Force (**Fenwick et al. 2020**) makes several 'good practice' recommendations. These recommendations support the inclusion of all uncertain parameters in the probabilistic analysis underpinning the value of information calculations. They also recommend considering the decision relevance horizon, research design and cost, and size of the beneficiary population when using value of information analysis. The report also states

that, when possible, structural uncertainties should be quantified and included in value of information analysis.

Considerations from workshop

124.   A workshop was organised to explore the potential use of value of information analysis as a tool to help inform decisions about evidence generation. The task and finish group considered a potential framework developed as an extension to the concepts in DSU report, which proposed that EVPI was compared with expected net benefit gain or loss when deciding whether evidence generation is worthwhile. The group also discussed the considerations outlined in the Claxton et al. paper, the difference between EVPI, EVPPI and EVSI, and how the value of information compares with the likely real-world value of any evidence generated following a research recommendation.

125.   Members of the group commented that EVPI was a simple extension of probabilistic analysis that would be a practical and useful addition to quantify the consequences of making a suboptimal decision because of uncertainty. It was noted that because EVPI is always an upper bound on this value so is higher than the real-world value of evidence generation, it might be most useful for illustrating when further research is unlikely to be worthwhile. For example, if EVPI – the maximum value of further research – is close to zero, then a decision-making committee may choose not to consider evidence generation recommendations. Task and finish group members who have sat on decision-making committees indicated that they may approximate value of information analysis in their heads when considering research recommendations, and that the presentation of EVPI analysis at committee meetings would make that process more transparent.

126.   Task and finish group members agreed that, although EVSI would reflect the real-world value of evidence generation better than EVPI, there are substantial barriers to implementing the analysis in current NICE processes. EVPPI was discussed as a potential compromise between EVPI and EVSI, but members considered that EVPPI may be too resource intensive when considering multiple scenarios, particularly in complex decision problems (for example, many treatment sequences).

127.   The task and finish group recognised that value of information analysis could only quantify parameter uncertainty, unless structural uncertainty is parameterised as recommended by Fenwick et al. As discussed earlier, this is likely to be challenging for many potentially plausible scenarios (see paragraphs 79 to 82). The group considered that this could be problematic in decisions in which structural uncertainty is a key component of the overall uncertainty.

128.   The task and finish group considered that there were many factors that decision makers would need to consider if presented with a value of information analysis. These include the considerations outlined by Claxton et al. and in the Grimm et al. DSU report. Members were keen to explore the implications of using value of information analysis on committee decision making. The suggested approach was to develop case studies and explore how the availability of value of information analysis would affect decisions. Task and finish group members noted that these case studies should reflect technologies likely to be encountered in CHTE in the future to ensure that the methods guide is 'future-proof' Members also suggested that further exploratory work should look at the feasibility of using EVSI in decision making because this is the measure most likely to reflect the real-world value of evidence generation.

129.   The specification of this task and finish group was to consider whether value of information analysis could be used when deciding if evidence generation could reduce uncertainty and increase the likelihood of making an 'optimal' decision. In this context, and because NICE does not fund evidence generation, the cost of generating further evidence was not included in the framework explored at the workshop. However, task and finish group members considered that the cost of evidence generation should be considered by decision makers. This is aligned with the recommendation by Claxton et al. that decision makers using value of information analysis should consider the likelihood of research taking place (because high costs might make research less likely to happen).

Cases for change to CHTE methods

130.   There was no consensus about the inclusion of decision rules or a formalised framework to incorporate value of information analysis into the CHTE methods guide. There was agreement that EVPI is easy to implement as an extension to probabilistic analysis; greater use of probabilistic modelling is recommended as a case for change in this update (see paragraph 20). Value of information analysis is already included as an option in the current DAP and MTEP methods but is not included in TA and HST methods.

131.   There was some support from the methods working group for proceeding to consultation with the proposal of routinely presenting EVPI to committees. While recognising its limitations, it was generally seen as a simple extension to probabilistic analysis that may provide committees with additional information that may be useful (that is, the maximum expected cost of current decision uncertainty). This is captured in the case for change text in paragraph 20.

132.   Further work and information are needed to assess whether a more comprehensive framework on the use of value of information analysis could be a potential case for change. Further research is needed to evaluate the likely implications of using value of information analysis as the basis for making research recommendations. The DSU report contains 4 retrospective case studies that apply value of information analysis to previous technologies, which could be used as a starting point for further research. Technology appraisal topics that have led to Cancer Drugs Fund recommendations could also provide useful case studies. Additionally, it is proposed that case studies based on forward-looking examples of likely future Health Technology Assessments should be used to explore the effect of value of information analysis. For each case study, decision makers could be presented with a value of information analysis based on the evidence used in the assessment. Decision makers could then indicate whether the analysis would have affected the recommendation(s) made.

Presenting uncertainty

133.   Various methods exist in Health Technology Assessments (HTAs) for presenting parameter uncertainty. For clinical inputs, confidence intervals and forest plots are commonplace. For model outputs, one-way sensitivity analyses are routinely presented as 'tornado' histograms, showing the inputs that have the biggest effect on cost effectiveness. Probabilistic model results are often presented on cost-effectiveness planes as scatter plots, showing each model simulation and the mean result. These can include the decision-making threshold value as a line, with the proportion of simulations below the line indicating the probability that the technology is cost effective compared with its comparator. Cost-effectiveness acceptability curves (CEAC) and frontiers (CEAFs) present this probability at different threshold values. However, there is no common method used to present structural uncertainty. Instead, it is typically shown as a series of tabular results to compare the cost-effectiveness estimates between alternative scenarios.

Current Centre for Health Technology Evaluation (CHTE) methods

134.   All methods guides across CHTE stipulate that uncertainty should be explored and appropriately captured in the analyses. The Technology Appraisal (TA) Programme guide is the most prescriptive about the methods that may be used to present uncertainty, citing most of the approaches described above. None of the guides explicitly needs an overall summary of uncertainty to be presented, although the Medical Technologies Evaluation Programme (MTEP) guide perhaps comes closest by needing its committee to explicitly 'describe the degree of uncertainty associated with [its] recommendations, and the potential impact of such uncertainties'.

Overall assessment of uncertainty

135.   Decision-making committees in CHTE are typically familiar with the various presentation techniques used to characterise parameter uncertainty. However, similar techniques are not routinely used to depict other types of uncertainty, which risks those uncertainties being given less attention in decision making.

Developments from policy, academia and other HTA bodies

136.   The Treasury's **Aqua Book (2015)** discusses 'deep' uncertainties, describing them as uncertainties that cannot be quantified or those for which nothing can be said about their effect on the decision outcome. The Book stipulates that the nature and causes of such uncertainties should be brought to the attention of decision makers.

137. **Grimm et al. (2019)** have developed a tool for analysts to systematically identify, assess and report uncertainty in decision models, to make their effect on cost-effectiveness estimates more transparent. The TRUST tool has 3 stages: formally identifying and categorising uncertainties; appraising how well they have been characterised in probabilistic and scenario analyses; and then using that overall assessment of uncertainty to inform decision making.

Cases for change to CHTE methods

138. For input parameters and structural uncertainties, usually a plausible range of values or preferred scenario can be informed by some form of evidence, such as expert advice. However, this might not be sufficient when there is unresolvable uncertainty. For example, in a rare-disease area with severe unmet need, trials might be small because of a very small patient population, single arm if no licensed treatment exists, or subject to treatment switching ('crossover') if an Independent Data Monitoring Committee mandates it for ethical reasons. These would introduce inherent uncertainty in the data, and might not inform plausible upper and lower bounds on its effectiveness. Such factors could be presented in an overall assessment of uncertainty to inform decision making, for example, using the TRUST tool. However, the CHTE methods guide should not be prescriptive about which tool should be used.

139. The following text is proposed for the updated CHTE methods guide:

An overall assessment of uncertainty should be presented to committees to inform decision making. This should describe the relative effect of different types of uncertainty (for example, parameter, structural) on cost-effectiveness estimates, and an assessment of whether the uncertainties that can be included in the analyses have been adequately captured. It should also highlight the presence of uncertainties that are unlikely to be reduced by further evidence or expert input.

140. Academic groups currently consider the various uncertainties in their critique of the evidence. However, the proposed change may need a new, dedicated section to be routinely written in their reports, so the overall assessment of uncertainty is clear and consistent between NICE assessments. Writing this summary would need additional work from the academic groups. Alternatively, NICE technical staff could produce this summary as a slide routinely presented to committees, extracting the relevant information from academic groups' reports, and liaising with them when necessary.

Net benefit rankings

141. CEACs (and CEAFs) are commonly used to present the results of probabilistic cost-effectiveness analyses. They show the proportion of probabilistic results that were below a given cost-effectiveness threshold value, which represents the probability that a technology is cost effective (optimal) at that threshold. They can include any number of comparators and threshold values.

142. However, CEACs are not routinely presented to committees across CHTE. They can become cluttered and difficult to interpret when several options are being compared, such as in a multiple TA. They typically show only the probability that each option is optimal, which may conceal uncertainty; for example, the clinical evidence for a technology might be so uncertain that it has a high probability of being cost effective, but at the same time also has a high probability of being the least cost effective. This may also conceal which technology has the highest expected net benefit (that is, most cost effective, or optimal), which might not be the same as the technology with the highest probability of being optimal. Further, NICE committees typically use 1 cost-effectiveness threshold (or a narrow range) when making a decision, meaning most of the thresholds used to generate a continuous CEAC are not informative.

Developments from policy, academia and other HTA bodies

143. No international HTA methods guidance was identified that asks for methods for presenting uncertainty that extend beyond the common approaches listed above. Methods guidance only varies by which methods they specifically need; for example, tornado diagrams, cost-effectiveness planes and CEACs are each included in about half of the methods guides reviewed. The least-cited presentation method was the use of a confidence ellipse to impose a confidence interval onto a scatter plot of probabilistic model results.

144. **Epstein (2019)** proposes an alternative to the CEAC, adapting a technique that is often used to present network meta-analysis (NMA) results for clinical evidence. It involves ranking treatments in each probabilistic simulation from best (highest net benefit) to worst (lowest net benefit) at a given threshold value, then, across all simulations, calculating the probability that each treatment ranks best and is therefore cost effective. This approach also allows the probability that each option is second best, third best, and so on to be calculated, which is not typically done using CEACs. In a decision space with several options, this approach presents the likelihood (or risk) of recommending a cost-ineffective option, but further, it also shows the risk of recommending a technology that is among the lowest-ranked options (that is,

very cost ineffective, imposing higher opportunity costs on the NHS). The author explains that these results can be presented as histograms, or 'rankograms', which may provide a simple visualisation.

145.   Two examples of rankograms are presented below, based on simulated probabilistic cost-effectiveness results comparing 4 possible options. In Figure 1, Option C is estimated to give the highest expected net (monetary) benefit (£2,406 per person) and has the highest probability of doing so (61%). Option B has the lowest expected net benefit and a high likelihood of having the lowest net benefit. In this scenario, a decision-making committee might be reasonably comfortable in recommending Option C. In Figure 2, the clinical evidence for Option C is subject to much greater uncertainty. It is still expected to provide the highest net benefit (£1,891 per person) and has the highest likelihood of doing so (46%). However, it also has a relatively high likelihood of providing the lowest net benefit (27%) because of considerable uncertainty about its effectiveness. The expected net benefit of Option A is very close to that of Option B, and A is highly likely to provide at least the second-highest net benefit (88%). Seeing this information might help a committee to better consider the risks posed by uncertainty in its decision making.

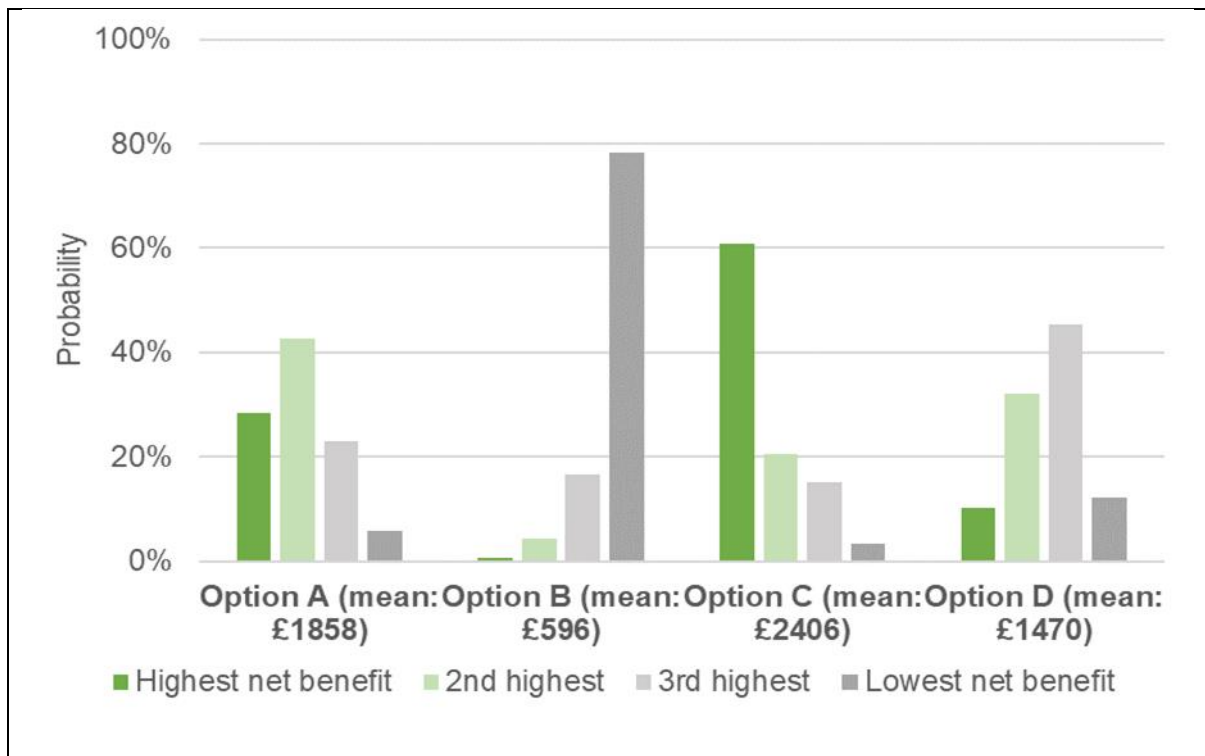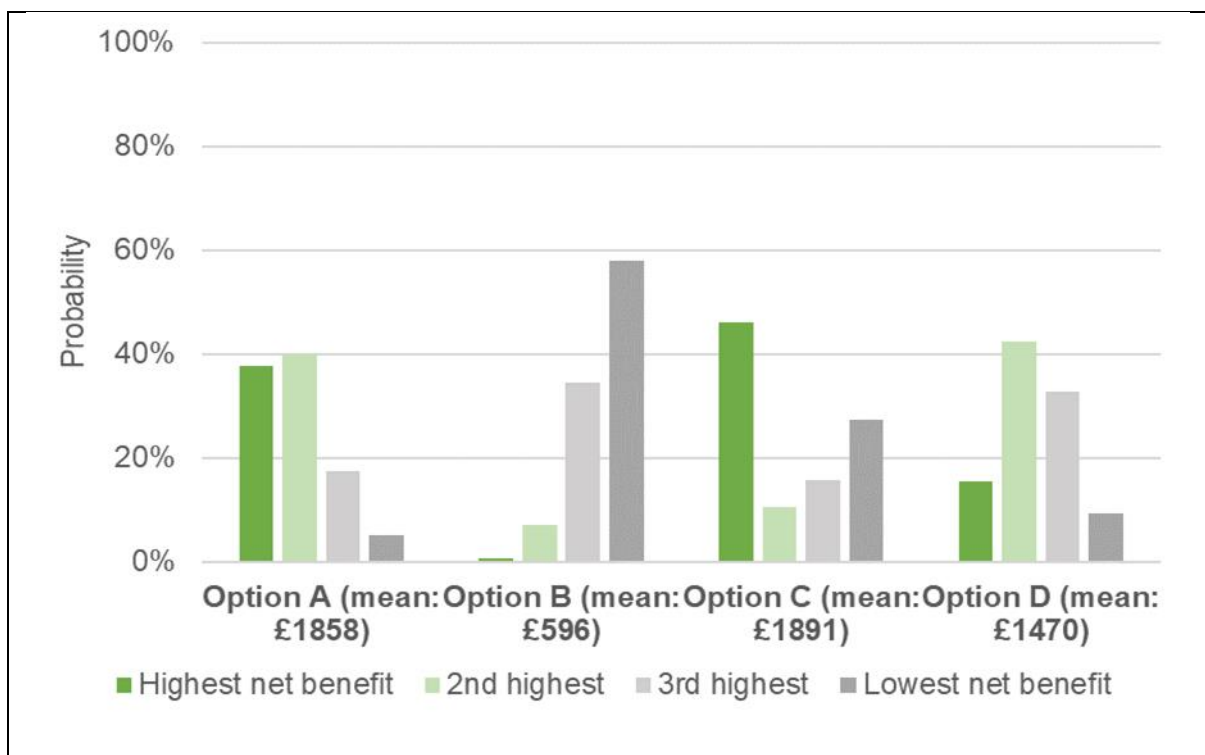**Figure 1 Example rankograms with relatively little uncertainty**

Figure 2 Example rankograms with relatively high uncertainty

Cases for change to CHTE methods

146.   Net benefit rankings, which can be presented as rankograms, may be useful for decision making when there are multiple technologies under consideration. For example, committees may be more acutely aware of the potential consequences of decision uncertainty if they see that a technology with a reasonable probability of being the optimal (most cost effective) option also has a notable probability of being the least cost-effective option, as well as the expected cost effectiveness. Rankings are less likely to be informative for single technology decision problems, when there will often be just 1 relevant comparator, or several comparators that are similarly effective.

147.   The following text is proposed for the updated CHTE methods guide:

When multiple technologies are being compared, cost-effectiveness rankings may be used to present the results of probabilistic model analyses. This should show the probability that each technology is ranked highest (produces the highest net benefit). It may also be informative to know the probability that each technology is ranked second, last, and all positions in between. Ranking-based histograms ('rankograms') may be used to present this information in a simple way, alongside the expected net benefit of each technology.

148.   The proposed text is not likely to cause implementation issues. Firstly, it does not stipulate that ranking-based approaches must be used, rather that they may be informative. They can easily be developed using the outputs of a probabilistic NMA or cost-effectiveness model. So presenting them is unlikely to place significant additional burden on the academic groups who typically conduct analyses with multiple technologies. Currently, when ranking-based results are presented, it is often a large table. A graph-based approach may not only be easier to understand, but also quicker to put together and edit than an unwieldy table.

149.   The methods working group noted that. when results are presented as rankings, this provides less information about the committee's preferred cost-effectiveness estimates, so may be less helpful for informing subsequent commercial discussions. To ensure the actual cost-effectiveness estimates are also presented alongside rankings, the proposed text includes the stipulation that rankings should be presented "alongside the expected net benefit of each technology".

Learnings from other (non-HTA) industries

150.   Our review of other HTA bodies' methods guides, key academic literature and Treasury guidance yielded little novel methodology for presenting uncertainty. With recent advances in data visualisation techniques, we had anticipated that this would be a rich area of research. This appears not to be the case. Therefore, industry members of the group

proposed and arranged a workshop, led by an external consultancy, to discuss presentation methods used to characterise and communicate uncertainty in other, non-HTA sectors.

151.   The workshop included presentational methods used in the aviation, insurance and utilities sectors. A key conclusion from the workshop was that CHTE methods do not appear to be missing a 'silver bullet' in how uncertainty is presented; other industries use similar techniques, and similar methods to propagate uncertainty in analyses (namely probabilistic, Monte Carlo simulations). There were 2 areas in which the task and finish group felt CHTE could learn from methods used in the aviation industry:

- categorising different types of uncertainty to improve how they are presented to committees

- visually showing when new evidence is expected to become available over time, including its potential effect on the level of uncertainty.

Categorising uncertainties by their resolvability

152.   The workshop discussed the use of colour coding in the aviation industry to categorise uncertainties by whether they are inherent or whether something can be done to mitigate them. Key uncertain inputs or assumptions that risk a negative outcome occurring are presented. They are highlighted blue if they are not considered to be practically resolvable, and green if a practical action can be taken to reduce the uncertainty and risk of a negative outcome (see Figure 3).

Figure 3. Colour codes used to present sources of uncertainty on OTP (on time performance) in the aviation industry
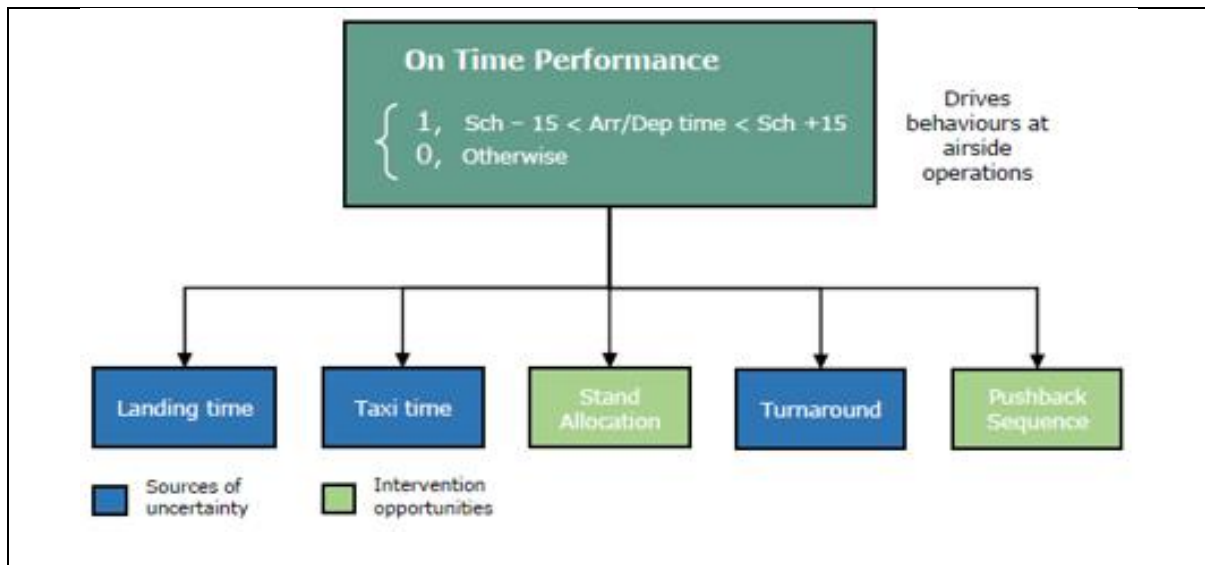

153. Adapting this to the context of HTA, an example of how this might look was presented by members of the group from industry, which has been adapted in Figure 4. Blue uncertainties, identified as not being practically resolvable in a useful time frame, could include things that a committee should be aware of that will not change during an appraisal or evaluation, or in the foreseeable future (for example, the licensed dose of a technology has recently changed since its trial was conducted). Green uncertainties would include factors for which something could reasonably be done to reduce or resolve the uncertainty. For example, uncertainty about whether the treatment pathway in a model reflects NHS practice could be informed by further expert elicitation, and uncertainty about how long the treatment effectiveness is sustained for could be informed by a future trial data cut. Clinically plausible ranges should be identified for green uncertainties, ideally as early as possible in the process, and cost-effectiveness estimates across the range should be presented for discussion, to allow the committee to conclude on its preferred, most plausible estimate. The group considered that most CHTE appraisals and evaluations also include a third category, not shown in Figure 4, of uncertainties that were adequately explored and addressed during the evidence critique stage, and therefore need less committee discussion.

Figure 4. Example figure showing colour-coded uncertainties
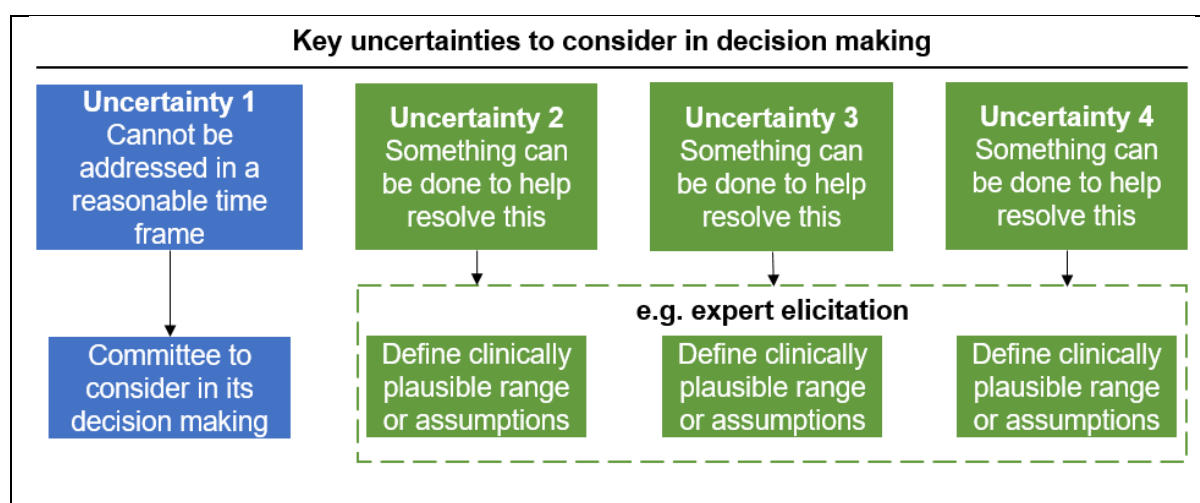
154.   Categorising uncertainties in this way may provide a simple method of focusing a NICE committee's discussions. For this to happen, the key uncertainties and the appropriate categories to use should be discussed and agreed before the committee stage (for example, during the clarification stage and, in TA, technical engagement). As expert elicitation may be needed to inform this, it may link with the request for the role of expert elicitation to be considered in the process review (see paragraph 43). The group agreed that the same colours could also be used in other, more common presentation techniques, such as tornado diagrams which are often used to show the most important model inputs (see **Erreur ! Source du renvoi introuvable.**).

Figure 5. Example colour-coded tornado diagram
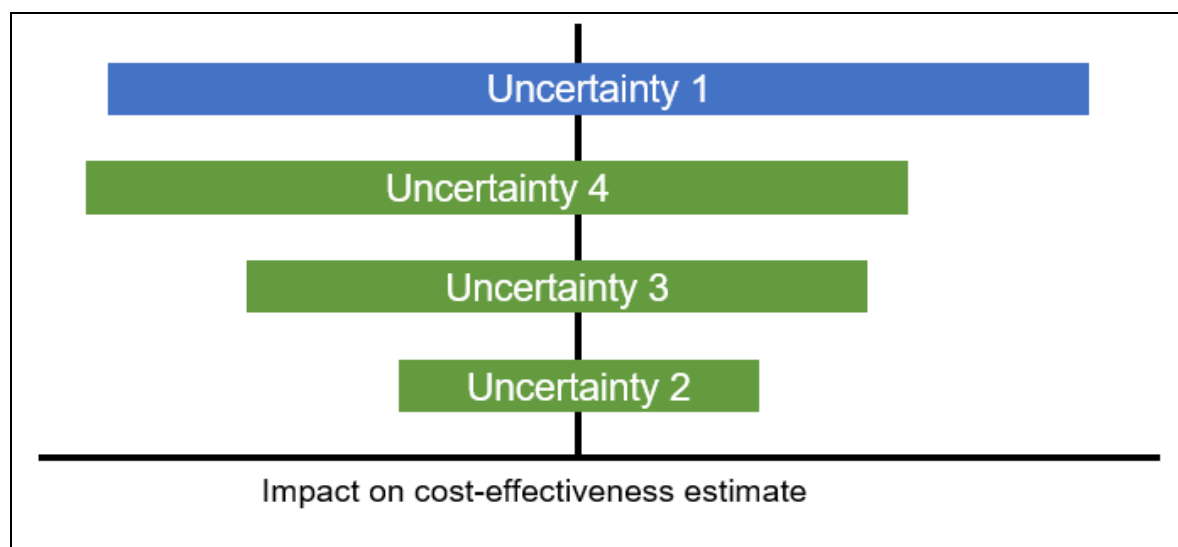
155.   Categorising different uncertainties might also have benefits in communicating the committee's decision, particularly in showing how they reached their preferred, most plausible cost-effectiveness estimate (or

range). Patient expert members of the task and finish group advised that this would make committee discussions clearer and more transparent. Industry members of the group advised that it would provide greater clarity in considering when commercial arrangements and further data may be more beneficial.

156.   A potential limitation of categorising uncertainties, raised at the workshop, was that defining different uncertainties as unresolvable or actionable may cause disagreements during the process. For example, a company might consider that a source of uncertainty is inherent, but the academic group might feel that it can practically be informed by additional evidence. Different clinical and patient experts might also provide different views on how resolvable an uncertainty may be, and a step of formal expert elicitation is not currently part of CHTE processes (see paragraph 44). There is therefore a risk that sorting sources of uncertainty into binary categories, intended to help focus committee deliberations, may itself need significant committee discussion.

157.   The methods working group agreed that the proposed approach may simplify uncertainties and make decision making more transparent. However, it felt that defining the importance and resolvability of different uncertainties is part of the committee's remit. If uncertainties are categorised before a committee meeting, this will lead to disagreements that will potentially need additional committee effort and time to resolve. Rather than categorising uncertainties to inform decision making, it may be more practical to do this after a committee meeting to help explain how the decision was reached in a transparent way.

158.   The working group also advised that the categories and wording should both be considered carefully. Simple 'resolved', 'resolvable' and 'unresolvable' groups may lack subtlety and objectivity. It may be difficult to consider an uncertainty 'unresolvable' as there might always be further evidence that would, theoretically, provide more information on uncertain aspects. Further, an uncertainty might be 'resolvable' if additional data are collected over 10 years, but that might be impractical in the context of a particular decision problem being considered by NICE. An example of a more nuanced approach might look something like the following:

a)   Uncertainties that were suitably accounted for during the assessment of the evidence. Not expected to need substantial committee discussion.

b)   Uncertainties that might plausibly be reduced by additional evidence in a reasonable time frame. Expected to need committee discussion.

c) Uncertainties that are highly unlikely to be reduced by additional evidence in a reasonable time frame. Expected to factor into committee decision making.

Cases for change to CHTE methods

159.   The working group advised that the approach of categorising different types of uncertainty and presenting them in a simple visual framework has merit. The approach used in the aviation industry is interesting and potentially adaptable to the HTA setting. The dual need to focus committee discussions on key uncertainties when possible, and to transparently present committee decisions, highlights a reasonable case for change to CHTE methods. However, further work and thought needs to be given to when the categorisation should happen and who does it. Consideration should also be given to exactly what the categorisation should look like, as a system that uses only colours to distinguish between different uncertainties would not be accessible to all people. The working group would welcome suggestions as part of the consultation process.

Presenting when future evidence will become available

160.   The workshop discussed the aviation industry's use of figures showing how future outcomes become increasingly uncertain, but a decision still must be made at the current point in time (see **Erreur ! Source du renvoi introuvable.**).

Figure 5. Expected effect of uncertainty on OTP (on time performance) in the aviation industry
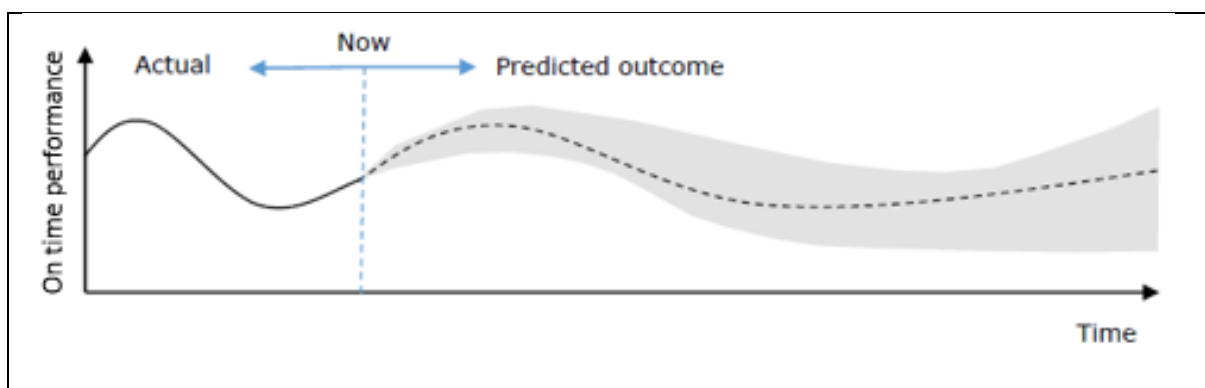
161.   In the context of HTA, this kind of figure could be adapted to show how uncertainty might be reduced over time with the arrival of new evidence. Committees across CHTE are already accustomed to thinking about the extent of ongoing research, and the plausibility of further research happening, in their decision making. An example was presented by members of the group from industry, which has since been developed further below (see Figure 6). It indicates the existing trial evidence, the current decision point,

time points at which evidence is expected to become available in the future, and the key parameters it should inform or uncertainties it may resolve.

Figure 6. Example figure showing expected evidence generation over time
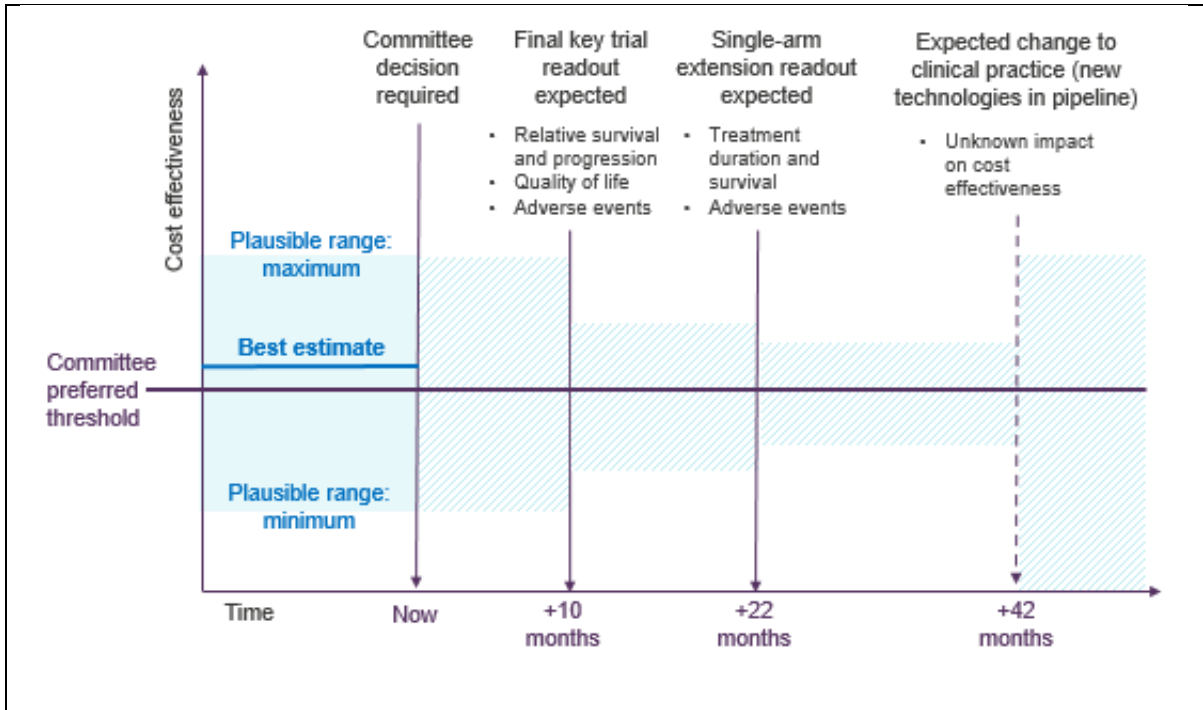
162. The task and finish group considered that a consistent way of presenting this graphically might help committee members place their current decision in the context of future evidence collection and allow committees to consider their recommendation options more openly, particularly in terms of time. For example, it might help focus committees on making a routine commissioning decision if no further data are likely to become available to resolve the existing decision uncertainty. Alternatively, if a trial readout is expected in 10 months, this may focus committees on the option of making a type of recommendation that formally recognises the ongoing data collection. This is likely to be most applicable to technologies with future evidence in the pipeline, or when evidence-generating recommendations are more likely.

163. The following text is proposed for the updated CHTE methods guide:

A graphical presentation of the evidence generation process for a technology over time, including planned future evidence generation, can be included in submissions to CHTE. This should show the expected time points of interim and final data readouts from ongoing clinical studies and planned additional studies. It should also indicate the key sources of uncertainty that might be reduced at each evidence-generating time point; for example, a forthcoming readout for a clinical trial may inform all aspects of relative effectiveness, while a future single-arm extension study may inform long-term survival outcomes for the technology.

164.   This could be included in company submissions to CHTE, for presentation to committees. This would support committees in considering the value of evidence generation recommendations, which they already routinely do but typically without a visual aid to help.

Survey of committee members

165.   To further supplement the group's discussions about how uncertainty is presented, we sought to take a retrospective assessment of how useful CHTE committee members find the existing, common presentation methods in their decision making. A brief survey was circulated to committee members asking whether they were familiar with:

- net benefits

- confidence intervals

- the 'probability cost effective' statistic (that is, the proportion of probabilistic cost-effectiveness estimates below the decision-making threshold)

- CEACs

- cost-effectiveness planes

- confidence ellipses

- tornado diagrams

- value of information analysis.

If the respondent was familiar with a method, they were asked to describe how useful they found it for decision making. Respondents were then invited to provide any further thoughts, for example to explain alternative ways they would like to see uncertainty presented.

166.   Most methods that committee members had experience of were described as 'very useful' or 'quite useful'. Numeric representations of uncertainty (for example, confidence intervals around incremental cost-effectiveness results) were reported as useful by 92% of responses, and 91% found tornado diagrams informative. Around 80% of responses said cost-effectiveness planes, CEACs, 'probability cost effective' results and net benefit results are useful for decision making. These approaches had all been seen in a committee setting by at least 76% of respondents. Among them, the response 'not at all useful' was only provided once, for net benefit results.

167.　　The 2 least-used methods were confidence ellipses (53% of respondents) and value of information analysis (38%). These were also considered to be the least useful for decision making. Only 55% of responses considered confidence ellipses to be at least 'quite useful'. This may, in part, be because of how well they are explained, given they aim to place a confidence interval on a cost-effectiveness plane, and numeric confidence intervals were reported to be very useful. Value of information analyses were reported as useful by just 31% of responses.

168.　　Committee members provided 18 free text responses. The most prominent theme (7 of 18) was a desire to pay greater attention to uncertainty by being presented probabilistic results more often, particularly the 'probability cost effective' statistic, confidence intervals and convergence of cost-effectiveness estimates, and the spread of model results on cost-effectiveness planes.

169.　　The second-most raised theme was how readily uncertainty results are understood (6 of 18). One respondent suggested that committee members would develop familiarity if a small number of presentation methods were needed consistently in all appraisals or evaluations. Other comments included recommending that a narrative summary of the overall uncertainty is presented; noting that wider (non-parameter) uncertainty remains important for decision making; that infographics may be useful for presenting complex data; and that results could be presented in more clinically meaningful ways (for example, confidence intervals for a 'number needed to treat' statistic).

170.　　Three responses suggested that cost-effectiveness estimates presented as net benefits would often be more useful than incremental cost-effectiveness ratios. Further comments stated that value of information analysis may help guide further evidence generation, particularly for individual parameters (3 of 18). However, using value of information analysis in a committee setting was described as challenging, and it was suggested that it might not ultimately convey more useful information than a tornado diagram.

Cases for change to CHTE methods

171.　　An overarching message from the survey results appears to be that uncertainty analyses are considered to be a useful part of the decision-making process. This supports our proposed cases for change which, together, seek to increase the extent to which decision uncertainty is considered by committees across CHTE.

172.　　That over 80% of respondents find probabilistic analyses informative supports our proposal to make probabilistic results the primary results for

decision making, with specific reference to cost-effectiveness planes, confidence intervals and the 'probability cost effective' statistic (see paragraph 20). These approaches have been mentioned to promote their routine use across CHTE and increase committee members' familiarity with understanding and interpreting them.

173. The request for a narrative summary of the overall uncertainty in a decision problem has been captured by our proposed change in paragraph 139.

174. The reported usefulness of tornado diagrams supports our proposal to retain deterministic analysis in CHTE methods, and warrants specifying that tornado diagrams may be used to present such results (see paragraph 20). They are not currently included in the 3 CHTE methods guide but are commonly included by international HTA bodies.

175. There is also a consensus that net benefit results are acceptable and may be preferred, which supports their inclusion in our recommendation about multiple technology decision problems (see paragraph 147).

176. A notable omission from the survey is threshold analysis; however, anecdotally, NICE staff members in the task and finish group are aware that it has been useful for decision making in previous assessments (particularly in DAP).

Equalities

177.   The task and finish group considered whether there were any potential equalities issues about exploring uncertainty and arising from any of the proposed new methods. The issue that was most prevalent was that there is likely to be more uncertainty in decision problems for rare and ultra-rare conditions. This is because rare conditions are generally associated with greater limitations in the clinical evidence base and therefore greater uncertainty in the evaluation.

178.   Many rare conditions may be classified as a disability, in addition to a number of rare conditions starting in children. Both age and disability are protected characteristics under the Equality Act 2010. The group considered the effect of the proposed methods on these groups. Two areas were highlighted that may have an effect.

179.   The first area was the increase in necessity for probabilistic analysis, given that there are greater limitations in the evidence base in rare conditions. It was noted that the primary purpose of using probabilistic analysis – to get the best available estimate of cost effectiveness – does not necessarily affect rare-disease technologies any more than others. Mechanisms that do not rely on data (namely, expert elicitation) exist that can be used to inform a probabilistic analysis and ensure the base-case result is as accurate as possible. Further, the alternative approach of using only deterministic results makes the strong assumption that we are certain about the values of input parameters. It was agreed that, in terms of characterising uncertainty – the main remit of this group – the proposal does not cause an unintended equalities issue in this regard. However, it is acknowledged that how uncertainty and rarity are used as a modifier in decision making could potentially create an equality issue.

180.   The second area was about the use of value of information analysis because rare conditions are more likely to benefit from evidence generation recommendations, given their inherent limitations in clinical evidence. The current proposal to need an Expected Value of Perfect Information will provide more information to committees, and potentially improve evidence generation recommendations. This will continue to be a consideration as further work is completed in this area.

Summary

181.   This report presents the task and finish group's proposed cases for change to methods across the Centre for Health Technology Evaluation (CHTE) for exploring uncertainty, partly informed by a survey of current committee members. The most notable suggestion is that probabilistic cost-effectiveness estimates should routinely be used in committee decision making whenever practical. This will need committees and reviewers to pay closer attention to the appropriateness of probabilistic analyses and the extent of parameter uncertainty. To support this, the group suggests incremental improvements to its methods about the parameterisation of ordered parameters, the plausibility of inputs and showing model convergence. It may also be practical to explore probabilistic one-way sensitivity analyses, and to incorporate alternative structural assumptions into probabilistic analyses. The group suggests these may be explored if given careful consideration.

182.   The group recognises that scenario and deterministic analyses remain useful tools to support decision making. Scenario analyses may be particularly informative when the evidence base is less robust, such as in very rare therapeutic areas, although it remains good practice for each scenario analysis to be probabilistic whenever possible. The group notes that correlation between parameters poses a challenge when conducting deterministic analyses, and it should always be considered carefully. A change is proposed to encourage the use of threshold analysis, when appropriate, for input parameters that lack robust evidence.

183.   The group has proposed cases for change about considering the plausibility of parameter values and scenario analyses. It notes that parameter values should be clinically plausible and derived using methods that are consistent with the data. Scenario analyses that explore the sensitivity of cost-effectiveness estimates to implausible scenarios are not useful for decision making, unless they show that results are robust to a source of uncertainty. It was agreed that it may be efficient to agree on the plausibility (or otherwise) of parameter ranges and scenarios early in an appraisal or evaluation using, for example, a framework similar to that used in the aviation industry (see paragraph 152). While companies should use formal expert elicitation to inform these in their submissions, further expert elicitation may be valuable. This does not currently happen in a formal way in CHTE. The group would like this to be considered as part of the process review.

184.   It is also recommended that CHTE methods need a detailed, transparent account of how a model was conceptualised, including why the

model structure was considered appropriate and detailing any involvement of experts. This will increase committees' confidence in underlying modelling decisions and assumptions. The group also makes specific recommendations about modelling beyond the available data, including the use of flexible methods, selection of appropriate extrapolations and capturing uncertainty introduced by adjustments for trial crossover.

185. The group did not reach a consensus about cases for change about the use of value of information analysis. It recognised that a lot of work and thought has gone into this topic before now, and some members explained that Expected Value of Perfect Information (EVPI) is a simple extension of probabilistic analysis that may be useful for decision making in some circumstances. However, others raised concerns that it may itself be subject to uncertain estimation (for example, in defining a relevant population or time horizon), and is much less informative for decision making than more involved approaches (for example, Expected Value of Sample Information). The methods review working group took the view that the use of EVPI, as an extension to probabilistic analysis and to provide additional information to committees, should be taken forward to consultation.

186. Proposals about how uncertainty is presented include needing a textual overall assessment of uncertainty to be presented to support decision making. It is also suggested that companies should consider providing a graphical depiction of ongoing and future evidence reporting times, to help committees consider the appropriateness of evidence generation recommendations. Further, net benefits and associated rankings are proposed as a way of presenting cost-effectiveness results of analyses with multiple technologies. A colour-coding system to transparently categorise uncertainties based on how resolvable they are, adapted from the aviation industry, was discussed. The working group agree that this poses an interesting case for change, but further work is needed to establish how and when the categorisation should happen, and what it should look like.

References

Annemans L, Makady A (2020) TRUST4RD: tool for reducing uncertainties in the evidence generation for specialised treatments for rare diseases. Orphanet Journal of Rare Diseases, 15(1): 127

Azfali H, Bojke L, Karnon J (2018) Model structuring for economic evaluations of new health technologies. Pharmacoeconomics, 36(11): 1309-19

Bell Gorrod H, Kearns B, Stevens J, et al. (2019) A review of survival analysis method used in NICE technology appraisals of cancer treatments: consistency, limitations and areas for improvement. Medical Decision Making, 39(8): 899-909

Bennett I, Paracha N, Abrams K, Ray J (2017) Accounting for uncertainty in decision analytic models using rank preserving structural failure time modelling: an application to parametric survival models. Value in Health, 21(1): 105-9

Bullement A, Latimer N, Bell Gorrod H (2019) Survival extrapolation in cancer immunotherapy: a validation-based case study. Value in Health, 22(3): 276-83

Claxton K, Palmer S, Longworth L, et al. (2016) A comprehensive algorithm for approval of health technologies with, without, or only in research: the key principles for informing coverage decisions. Value in Health, 19(6): 885-91

Epstein D (2019) Beyond the cost-effectiveness acceptability curve: the appropriateness of rank probabilities for presenting the results of economic evaluation in multiple technology appraisal. Health Economics, 28(6): 801-7

EUnetHTA (2015) Methods for health economic evaluations: a guideline based on current practices in Europe [online; accessed 30 October 2019]

Fenwick E, Steuten L, Knies S, et al. (2020) Value of information analysis for research decisions—an introduction: report 1 of the ISPOR Value of Information Analysis Emerging Good Practices Task Force. Value in Health, 23(2): 139-50

Grimm S, Pouwels S, Ramaekers B, et al. (2019) Development and validation of the TRansparent Uncertainty ASsessmenT (TRUST) Tool for assessing uncertainties in health economic decision models. Pharmacoeconomics, 38(2): 205-16

Grimm et al. (2016) NICE Decision Support Unit report. Framework for analysing risk in health technology assessments and its application to managed entry agreements, University of Sheffield

Guyot P, Ades A, Beasley M, et al. (2017) Extrapolation of survival curves from cancer trials using external information. Medical Decision Making, 37(4): 353-66

Hatswell A, Bullement A, Briggs A, et al. (2018) Probabilistic sensitivity analysis in cost-effectiveness models: determining model convergence in cohort models. Pharmacoeconomics, 36(12): 1421-6

HM Treasury (2015) The Aqua Book: guidance on producing quality analysis for government [online; accessed 6 November 2019]

HM Treasury (2018) The Green Book: central government guidance on appraisal and evaluation [online; accessed 6 November 2019]

Jackson C, Stevens J, Ren S, et al. (2017) Extrapolating survival from randomised trials using external data: a review of methods. Medical Decision Making, 37(4): 377-90

Kaltenthaler E, Tappenden P, Paisley S, Squires H. (2011). NICE Decision Support Unit Technical Support Document 13: identifying and reviewing evidence to inform the conceptualisation and population of cost-effectiveness models

Kearns B, Stevens J, Ren S, Brennan A (2020) How uncertain is the survival extrapolation? A study of the impact of different parametric survival models on extrapolated uncertainty about hazard functions, lifetime mean survival and cost effectiveness. Pharmacoeconomics, 38(2): 193-204

Kim D, Guzauskas G, Bennette C, et al. (2020) Influence of modelling choices on value of information analysis: an empirical analysis from a real-world experiment. Pharmacoeconomics, 38(6): 171-9

Kunst N, Wilson E, Alarid-Escudero F, et al. (2020) Computing the expected value of sample information efficiently: practical guidance and recommendations for four model-based methods. Value in Health, 23(6): 734-42

Latimer N (2011, 2013) NICE Decision Support Unit Technical Support Document 14: survival analysis for economic evaluations alongside clinical trials – extrapolations with patient-level data

McCabe C, Paulden M, Awotwe I, et al. (2020) One-way sensitivity analysis for probabilistic cost-effectiveness analysis: conditional expected incremental net benefit. Pharmacoeconomics, 38(2): 135-41

Othus M, Bansal A, Koepl L, et al. (2017) Accounting for cure patients in cost-effectiveness analysis. Value in Health, 20(4): 705-9

Ouwens M, Mukhopadhyay P, Zhang Y, et al. (2019) Estimating lifetime benefits associated with immune-oncology therapies: challenges and approaches for overall survival extrapolations. Pharmacoeconomics, 37(9): 1129-38

Ren S, Minton J, Whyte S, et al. (2018) A new approach for sampling ordered parameters in a probabilistic sensitivity analysis. Pharmacoeconomics, 36(3): 341-7

Roberts M, Russell L, Paltiel AD, et al. (2012) Conceptualizing a model: a report of the ISPOR-SMDM Modelling Good Research Practices task force-2. Value in Health, 15(6): 804-11

Sculpher M, Palmer S (2020) After 20 years of using economic evaluation, should NICE be considered a methods innovator? Pharmacoeconomics, 38(3): 247-57

Tappenden P (2014). Conceptual modelling for health economic model development. Health Economics and Decision Science discussion paper no. 12.05, University of Sheffield

Appendix

Summary of work completed compared with the tasks and questions specified in the project specification

Tasks	Questions	Outcome
1. Review NICE's existing work on handling uncertainty and explore how it might be incorporated into this project	1. What methods are used to mitigate, characterise, and present uncertainty in the Centre for Health Technology Evaluation (CHTE)? a. Are these methods being used appropriately and consistently?	**Complete** This question has been addressed throughout the report, with each section discussing 'current CHTE methods'. Each section identifies the similarities and differences between programmes.
2. Review how uncertainty is managed by other Health Technology Assessment (HTA) bodies and programmes	2. What methods are used to mitigate, characterise, and present uncertainty by other HTA bodies and actuarily? a. Are these methods applicable or practical for use within CHTE?	**Complete** This question has been addressed throughout the report, with each section discussing 'Developments from policy, academia and other HTA bodies'.

3. Review how parameter uncertainty is characterised, including:
    3a. The use of one-way, two-way and probabilistic sensitivity analysis
	3. What other methods or approaches exist to mitigate, characterise, and present uncertainty? These may include, but are not limited to:
	    3a. Sensitivity analysis
		Complete
		Sensitivity analysis is discussed in several places in the report, with the main focus on probabilistic sensitivity analysis and one-way sensitivity analysis (see paragraphs 10 to 25 and 46 to 52).

3b. The availability and practicality of using a good practice guide for probabilistic sensitivity analysis	3b. No specific question in the specification	**Partially complete** A specific good practice guide was not identified during the literature review. However, several good practice improvements were identified and have been identified as cases for change (see paragraphs 26 to 52).
3c. The use of threshold analysis	3c. Threshold analysis	**Complete** Threshold analysis was explored and a case for change has been identified (see paragraphs 53 to 62).

3d. The methods for extrapolating beyond observed data, such as: i. survival and progression-free survival ii. long-term relative effectiveness (for example, continued treatment effect) iii. long-term safety outcomes (adverse event rates) iv. time on treatment v. outcomes after stopping treatment	There was no specific question relating to extrapolation	**Partially complete** Extrapolating beyond the data is explored in the report and several cases for change have been identified (see paragraphs 84 to 112). The report considers extrapolating of time-to-event data and treatment benefits, flexible survival models, appropriately extrapolating and adjusting for treatment switching. No literature was identified on extrapolating adverse event rates so this has not been explicitly explored. However, some of the methods explored will be relevant.
No specific task for this question	3f. In all the above, consideration will be given to how the methods might be used in circumstances in which uncertainty is inherent and unavoidable, for example, in rare diseases.	**Complete** Consideration to inherent and unavoidable uncertainty has been explored throughout the task and finish groups work, with specific attention on rare diseases. The presentational work suggests approaches to distinguish between unavoidable and resolvable uncertainties. It was noted that it is very rare when no information would be known, and that expert elicitation should be used when there is an absence of quantitative data. Furthermore, it was noted that a high level of uncertainty is not a rationale to not do uncertainty analysis.
3e. Characterisation of parameter uncertainty arising from the use of surrogate outcomes	No associated question	**Not complete** This is not covered in this report as surrogate outcomes are now included in the evidence sources and synthesis task and finish group.
4. Review how structural uncertainty is mitigated and characterised, including: 4a. The use of scenario analysis	3b. Scenario analysis	**Complete** Scenario analysis has been explored and a few cases for change have been identified (see paragraphs 63 to 83).

4b. The use of model conceptualisation to mitigate structural uncertainty	3d. Model conceptualisation	**Complete** Model conceptualisation was considered and a case for change has been made (see paragraphs 66 to 73).
No direct task stated in the specification	4. What are the benefits and limitations of the different methods or approaches identified?	**Complete** Benefits and limitations are discussed throughout the report.
5. Review how uncertainty is presented to committees to support decision making and to stakeholders to support engagement and transparency 5a. Identify barriers to comprehension 5b. Consider methods to simplify or improve presentation, including ways to distinguish between clinical and cost uncertainties and innovative visual presentation	7. How can the outputs from these methods be presented to the committee? a. How do these outputs inform committee decision making? b. Are there any barriers to comprehension?	**Complete** The presentation of uncertainty and results from uncertainty analysis has been explored and a case for change has been identified (see paragraphs 133 to 176).
6. Review and consider the use of value of information analysis in HTA decision making 7. Identify methods for managing uncertainty that could potentially be used to inform managed access or research recommendations	3e. Value of information 8. Can the methods for managing uncertainty be used to inform evidence generation recommendations (including managed access or research recommendations)?	**Complete** Value of information analysis has been explored and a case for change has been identified (paragraph 113 to 132).
8. Review the practicality of aligning approaches to managing uncertainty across CHTE programmes (Diagnostics Assessment, Highly Specialised Technologies, Technology Appraisals and Medical Technologies Evaluation Programmes)	6. How can these methods be captured and addressed in the methods guide(s)?	**Complete** Each case for change has suggested revised wording to be adopted in the new methods guide.
9. Develop and recommend a framework for mitigating, characterising and presenting uncertainty, with consideration for NICE's	No specific question	**Partially complete** No specific framework has been developed, but it is considered that the cases for change alongside the already

commitment to the 2019 Voluntary Scheme		existing methodology provide a framework. The format in which this is incorporated into the final methods guide will guide if and how a framework is developed.
10. Assess the practical implications of implementing the recommended framework.	5. Can these methods or approaches be implemented pragmatically and consistently across CHTE?	**Complete** Implications of implementation have been explored for each case for change.
11. Consider potential equality implications for all options and proposals (when not otherwise covered by the Equalities task and finish group)	No specific question	**Complete** The main equality consideration has been made considering rare and ultra-rare conditions in light of our obligations not to discriminate against a protected characteristic, specifically disability. This has been considered in each section in which there was a potential implication and is summarised in paragraphs 177 to **Erreur ! Source du renvoi introuvable.**.

Authors

Project leads

Lucy Beggs
HTA adviser (TA)

Henry Edwards
Commercial risk assessment adviser

Jamie Elvidge
HTA adviser (TA)

Project team

Jasdeep Hayre
Associate director

Rufaro Kausi
HTA adviser (TA)

Tosin Oladapo
HTA analyst (DAP, MTEP)

Thomas Walker
HTA adviser (DAP)

Ian Watson
HTA adviser (lead)

on behalf of the exploring uncertainty Task and Finish group.