ECONOMIC EVALUATION IN NICE EARLY VALUE ASSESSMENTS

REPORT BY THE DECISION SUPPORT UNIT

November 2023

Authors: Mon Mon Yee[1], Paul Tappenden[1], Allan Wailoo[1]

[1] Sheffield Centre for Health and Related Research (SCHARR), University of Sheffield

ABOUT THE DECISION SUPPORT UNIT

The Decision Support Unit (DSU) External Assessment Centre is based at the University of Sheffield with members at York, Bristol, Leicester and the London School of Hygiene and Tropical Medicine. The DSU is commissioned by The National Institute for Health and Care Excellence (NICE) to provide a research and training resource to support the Institute's Centre for Health Technology Evaluation Programmes. Please see our website for further information www.nicedsu.org.uk.

The production of this document was funded by the National Institute for Health and Care Excellence (NICE) through its Decision Support Unit. The views, and any errors or omissions, expressed in this document are of the authors only. NICE may take account of part or all of this document if it considers it appropriate, but it is not bound to do so.

Acknowledgements

We would like to thank the following for helpful comments on previous drafts of this document: Matt Stevenson, John Cairns, Marta Soares, Bethany Shinkins, Janet Bouttell, Laura Trigg, Ed Wilson, Dawn Lee, Mary Jordan, Peter Auguste, Becky Albrow, Thomas Walker.

This report should be referenced as follows:

Yee, M., Tappenden, P., and Wailoo, A. NICE DSU Report. Economic Evaluation in NICE Early Value Assessments, 2023.

Contents

Tables

Figures

1 INTRODUCTION

In June 2022, the National Institute for Health and Care Excellence (NICE) launched its Early Value Assessment (EVA) programme. The EVA programme aims to improve the care of people and effective use of NHS resources through quicker access to and further evidence generation on promising digital products, devices and diagnostic technologies in areas where the evidence base is still in its infancy. EVA is intended to achieve these goals more quickly than would be the case through standard NICE appraisal routes. To support the launch of this programme, NICE has adopted a 'test and learn' approach which will involve at least 10 pilot case study EVAs. In December 2022, NICE published an interim statement which sets out the aims of the programme and the interim process and methods for EVAs.[1]

The interim statement[1] describes three aims of economic evaluation work carried out to inform EVAs: (i) to identify likely impacts of using technologies (whilst further data are collected); (ii) to identify additional uncertainties that would not be apparent from technology-related studies, for example, those related to the structure or parameters of a model required to inform future guidance and (iii) to identify uncertainties that are likely to be key drivers of model results and decision uncertainty to inform decision-making about further evidence generation. The interim statement recognises that economic analyses undertaken for EVAs may differ from standard approaches used to inform recommendations within other NICE appraisals (hereafter referred to as "full models") and that the economic evaluation work undertaken for EVAs that is likely to be most beneficial for committee decision-making may vary between topics. The interim statement indicates that efforts should be made by External Assessment Groups (EAGs) to identify relevant economic evaluations which could inform the economic analysis and that any models submitted by companies may also be considered. It also notes that other existing economic models could be used if considered suitable and if an agreement is in place to allow models developed by third parties to be made available to stakeholders. The interim statement is less prescriptive regarding whether and how any *de novo* economic model should be developed for use within an EVA. Whilst it is recognised that ideally, a preliminary or early coded model should be considered by EAGs to help meet the objectives of the EVA, this is not stated as a mandatory requirement of the process. The interim statement suggests that, at a minimum, the economic analysis should consider model structures (i.e., conceptual models) that would be needed for a future analysis to support NICE Appraisal Committee decision-making. If no coded model is produced, the interim statement states that searches should still be conducted to attempt to identify data for model parameters which are likely to be key drivers of future cost-utility analyses. The timescales for EVAs have thus far been limited to nine weeks for EAGs to undertake evidence reviews and any economic analyses. In a standard STA appraisal, EAGs have 6 weeks to perform a review of evidence submitted by

the manufacturer. For a diagnostic appraisal, the EAG have around 24 weeks to develop a final report.

As of September 2023, seven pilot EVA projects for medical and diagnostic technologies had been completed and published, and ten are still in development.[2] Amongst these, seven (out of eleven where the assessment report is in the public domain) EVA projects did not involve the development of a fully coded economic model (3/7 completed EVAs and 4/4 incomplete EVAs but where the assessment report is published on the NICE website). Within these projects, the reasons why models were not developed were because of limited data on the effectiveness of the technology and its downstream clinical consequences, limited data on costs, and structural uncertainty around the appropriate care pathway. In one of the case studies, data were available but not in the population of interest for the appraisal. In five out of these seven EVAs, model development was restricted to conceptual modelling only.

Owing to the early nature of EVAs, the key evidence on clinical effectiveness or diagnostic performance available to inform the appraisal is likely to be limited and may still be under development. Where quantitative economic models are developed, these may be considered as “early models” due to inherent uncertainty in their parameters and/or their structure. There is an existing body of literature around what early models are, what they involve and what they can be used for. NICE believes that EAGs may benefit from advice on what approaches to take for EVA assessments to make the most of the limited time available, and to reduce variability in what different EAGs would do if given the same project.

This report aims to provide an overview of the existing literature around the definition and value of early health economic modelling, as well as recommendations around how such models should be implemented and evaluated. The report also aims to explore analytical approaches and practical steps that EAGs could take when conducting economic analysis to inform EVAs, taking into account the likely information that would routinely be available in EVAs, time constraints and the decision-making requirements of NICE Appraisal Committees. In order to address these aims, we conducted targeted literature searches to identify existing articles on early economic modelling in health technology assessment (HTA) and reviewed the characteristics of the published pilot EVAs and any accompanying coded models developed to inform these appraisals. Based on the review findings, we then developed a general recommendation which puts forward general guidance which is intended to help EAGs decide on whether and how to develop early economic models to inform future EVAs.

The structure of this report is as follows: Section 2 describes the methods of the review of existing literature, the review of pilot EVAs and the generation of the recommendations for early modelling. Section 3 presents the findings of the review of literature and the review of pilot EVAs. Section 4 summarises the discussion around recommendations and other factors to consider when considering how to conduct early economic analyses alongside an EVA.

2 METHODS

Stage 1: Review of existing papers on early modelling/HTA

We conducted a targeted literature review to identify papers which discuss the value of early models and/or which provide recommendations on how to implement and/or evaluate such models. We also extracted information on how authors of these studies defined "early models" or "early HTA." As a starting point, we identified two key papers on early modelling, Scholte *et al.*, (2023)[3] and Grutters *et al.*, (2019),[4] together with a series of published responses to these papers. We hand-searched the reference lists of these papers and the responses to identify any further papers discussing the themes of early modelling and early HTA. We also performed additional electronic database searching using PubMed and Medline (via Ovid). The following keywords were used: "early value assessment", "early heath economic models", "early economic evaluation", "early HTA", "health technology", "medical technology" and "diagnostic technology." These additional database searches were conducted on the 11th August 2023. In order to be eligible for inclusion in the review, papers had to: (a) discuss the value of early modelling and/or (b) report recommendations for undertaking early modelling or analysing early models. Articles which described case studies of early modelling exercises, but which did not provide guidance or recommendations on how to implement such models, were excluded from the review.

Stage 2: Summary of economic analyses conducted in previous EVAs

As of 1st September 2023, seven EVAs have been published on the NICE website, and ten are still in development. We reviewed the available assessment reports and guidance documents for all published EVAs and the four ongoing EVAs. We extracted information relating to: the intervention; comparators; population; outcomes; the extent of the economic analysis conducted (full coded model versus conceptual model only) and any justification for this as well as details of methods used to evaluate the model (deterministic/probabilistic sensitivity analysis, scenario analysis and value of information analysis).

Stage 3: Recommendations for early modelling to support future EVAs

We make a series of proposals and issues for consideration for analysts required to develop reviews of evidence for NICE. These proposals set out how an economic analysis can be developed given the limited timeframe and resource available to do so, and the types of questions that such analyses can be expected to answer. The proposals draw on the literature review where relevant. These will form the basis of consultation with a range of stakeholders and used in a final version of this report.

3 RESULTS

3.1 Literature review findings

A total of 23 articles were selected for inclusion in the review based on the eligibility criteria described in Section 2. The findings of published papers describing early models are presented in three sub-sections. Section 3.1.1 provides definitions of early economic evaluation or similar terminologies used in the included papers. Section 3.1.2 describes the perceived value of developing early models. Section 3.1.3 provides a summary of recommendations provided in these papers, including key issues around conducting early modelling of diagnostics and medical technologies.

3.1.1 Definitions of early economic modelling

Based on a scoping review of early HTA, IJzerman *et al.*[5] states that the first papers presenting an early health economic model for medical devices were published in 2006. In the context of HTA, Vallejo-Torres *et al.* defines the "early stage" of an evaluation as being characterised by the limited availability of clinical and economic data, which means that it is challenging to apply standard health economic evaluation methods to inform decisions.[6] IJzerman *et al.*[5] define "early" as the situation in which there is uncertainty in the clinical evidence and the mechanisms to reduce or mitigate uncertainty in evidence development. Love-Koh[7] defines "early" as any point in time before healthcare payers are making decisions about whether to adopt the intervention. Markiewicz *et al.*[8] suggest a definition for the early assessment of medical devices as *"the assessment of the value of a new medical device at the time when the investment decisions are made with high uncertainties."* All of these definitions are broader than the NICE context of concern here. However, the high degree of uncertainty and the requirement to inform a decision that is considered interim are the key characteristics common to all thee definitions and of relevance to NICE.

IJzerman *et al.*[5] highlights that whilst the concept of early-stage health economic modelling has been around for some time, there are differences in definitions used to describe this

concept by different authors. They state that the most frequently used definition of early HTA is *"the use of economic evaluation in early stages of product development mainly to inform industry at the time that investment decisions are made."* IJzerman *et al*. proposes a new definition of early HTA as *"all methods used to inform industry and other stakeholders about the potential value of new medical products in development, including methods to quantify and manage uncertainty."*

Much of the literature on early modelling focuses on the perspective of the manufacturer and discusses the role and value of undertaking early HTA whilst the technology is still in development. This is likely to reflect a different context to NICE EVAs, as the timelines, purpose and resources available for early modelling may be very different. For the purposes of this report, the concept of "early modelling" is intended to broadly reflect the definitions provided in the literature, that is, economic modelling undertaken at a point where an adoption decision has not yet been made, where evidence generation on the primary clinical effect or diagnostic accuracy of the technology is still underway, and where the evidence that is currently available to demonstrate the value of the technology is extremely limited and therefore subject to considerable uncertainty, exceeding that which is typical for most , though not all, technologies at the point of a standard assessment.

3.1.2 The aims and value of early HTA and early modelling

Several of the papers included in the review discuss the aims and value of early economic modelling. Markiewicz *et al.*[8] state that early assessment aims not only to decrease the failure rate during each stage of product development, but also to enhance research and development (R&D) processes with resource constraints. Early modelling is intended to inform people involved in the R&D processes including manufacturers, investors and public funding bodies, regarding the information that could guide technology development. Early modelling aims to estimate whether and when (i.e., under what conditions) the new technology could potentially be cost-effective before further development is invested in.[9] In other words, the effectiveness, cost-effectiveness and safety of a new technology can be assessed given preliminary evidence and assumptions.[10, 11]

Vallejo-Torres *et al.*,[6] state that the early analysis of a new technology helps to prioritise the development of several potential devices when resources are limited.[6] As highlighted by Abel and Shinkins,[9] the cost-effectiveness estimates generated by an early economic model can later reduce the commercial risk associated with continued product development. Based on cost-effectiveness evidence, an innovative technology can be tailored according to clinical needs and care pathways, or discarded as early as possible if the model suggests that it is

unlikely to be cost-effective, depending on risk preferences. Abel and Shinkins also suggest that where the value of a technology has not yet been fully characterised, early modelling allows the flexibility to explore both interim and final outcomes in accordance with evidence needs at the time of the analysis.[9] Grutters *et al.*[4] and Grabowski[12] each suggest that early models can help to determine the indications for which the new technology should be targeted.[12] Grutters *et al.*[4] also comments that early modelling can help to identify sensitive parameters which are likely to drive the cost-effectiveness of the technology which can help to inform the design of future data collection.[4]

Some papers, including Abel and Shinkins *et al*.[9] and Fasterholdt *et al.*[13] state that an early model is more valuable when it is structured and implemented similar to a full model that can be updated later based on further evidence.

3.1.3 Summary of recommendations for developing and evaluating early economic models

Amongst the papers included in the review, several alternative economic evaluation approaches are mentioned. These include early assessments using cost-effectiveness and cost-utility analysis, as well the use of cost-consequence analyses and/or budget impact approaches.[8, 9] With respect to appropriate model structures for early models, Grutters *et al.*[4] suggest that decision tree or state-transition models are commonly used. Most of the models included in the scoping review by IJzerman *et al.*[5] and Fasterholdt *et al.*[13] used these approaches. These model structures are commonly developed and used to inform full NICE appraisals of diagnostic technologies.

As with full economic models, uncertainty may be analysed using a range of sensitivity analysis methods, including deterministic sensitivity analysis (DSA), scenario analysis, threshold analysis, probabilistic sensitivity analysis (PSA) and value of information (VOI) analysis. DSA allows for the identification of the main drivers of cost-effectiveness. PSA is used to estimate incremental cost-effectiveness ratios (ICERs) taking into account any non-linearities in the model, as well as providing estimates of the probability that a technology is cost-effective given uncertainty around all model parameters, and can be extended to include VOI analysis to determine the value of undertaking additional research and to prioritise further evidence development.[5] Of particular note, many of the papers included in the review around early health economic modelling refer to an analytical method call “headroom analysis.” This approach appears to be commonly used in early economic models and is discussed in several articles.[4, 9, 11, 14-16] Cosh *et al.*[15] describe the headroom method as a simple threshold analysis approach, whereby *“the headroom is the maximum net incremental cost for which the technology could be still cost-effective, and it is calculated based on the most optimistic*

assumptions in the plausible range of effectiveness data." E.A.Boudewijins *et al.*[16] conducted a systematic review of the headroom analysis in early economic evaluation and reviewed the application of headroom analysis in 42 early models of various types of interventions including diagnostic or screening tests, medicines, procedures, medical devices etc. Other authors provide similar descriptions of this method. This approach is particularly relevant from the perspective of the test/device developer as it provides an early estimate of the maximum price of the technology, given assumptions about its effectiveness. Among the early models included in the E.A.Boudewijins *et al*, ten percent of them estimated an effectiveness-seeking headroom instead of a cost-seeking headroom. The headroom analysis does not require a particular type of model or complexity, and it is a form of threshold analysis which can be applied to any type of model. Depending on the level of uncertainty in the effectiveness or performance of the intervention, headroom analysis might be the main analysis or a form of sensitivity analysis.

Across the range of papers included in the review, several authors offer suggestions and/or recommendations for structuring, parameterising and evaluating early models. These are summarised below across six main themes: (1) general process recommendations; (2) model structure; (3) model simplifications; (4) identification of evidence to inform model parameters; (5) use of expert opinion and (6) model evaluation methods.

3.1.3.1 General process recommendations for early model development

Abel and Shinkins,[9] outline five recommendations for researchers developing early models of diagnostic tests:

(1) Establish key questions to be answered. This should involve defining the purpose of the model to be addressed and the most appropriate type of economic analysis, which may be broader than other types of HTA (potentially including headroom analysis and budget impact analysis). Information gathering at this stage may help to identify risks and barriers to the adoption of the technology at an early stage in product development.

(2) Develop a model that reflects the care pathway. This stage relates to determining the level of granularity that should be reflected in the economic model, subject to the resources available for model development.

(3) Undertake early and frequent stakeholder engagement. This stage may relate to identifying key outcomes as well as ensuring that the modelled pathway with and without the diagnostic test is plausible, acceptable and generalisable. Relevant clinical stakeholder input may also be useful for ensuring the appropriateness of model inputs and the credibility of model results.

(4) Use adaptive review methods. Rapid review and iterative search methods are recommended over traditional systematic review search methods due to their improved

efficiency. These may be combined with stakeholder engagement to rapidly inform the structure and parameterisation of the model.

(5) Include meaningful sensitivity analysis. It is suggested that sensitivity analyses using plausible ranges of input parameter values are likely to be more useful than reporting a headline ICER.

3.1.3.2 Recommendations around developing the structure of early models

Abel and Shinkins[9] indicate a preference for developing comprehensive (full) models that can be updated as the technology develops, rather than adopting a "quick and dirty" approach. However, they also note that there is a balance to be struck between developing complex models which can be adapted at a later timepoint, and developing efficient, comparatively simpler models which can be used to answer specific questions quickly. With respect to the extent of the complexity of the early models, Abel and Shinkin[9] suggest that models for multiple care pathways (e.g., an imaging test which could be used for both screening for cancer and for surveillance of another unrelated condition) are useful but that they can be time-consuming to develop. The authors suggest that a pragmatic approach might involve implementing a simplistic decision tree analysis for each individual pathway, but they highlight that the validity of this is unclear. Hjelmgren *et al.*[17] suggest that where economic models have previously been developed within a given disease area, there may be advantages in an early model following the recommendations of other authors regarding various aspects of how that disease should be modelled (e.g., the time horizon, the selection of economic outcomes and approaches for modelling mortality and disease progression).[17, 18] Scholte *et al.*[3] discuss the concept of adding a new decision tree to an existing long-term model which is used to estimate the costs and health effects associated with particular branches. The reviews reported by Grutters *et al.*[4] and IJzerman *et al.*[5] indicate that many early economic models have adopted a decision tree approach, a state transition model, or a combination of the two. There are some exceptions to this, as one of the models included in the scoping review by IJzerman *et al.* adopted a systems dynamics approach.[19]

3.1.3.3 Recommendations around model simplifications

Some early models have included simplifications by focussing on intermediate rather than final endpoints as the metric of health benefit through which to estimate the effect of the new technology. For example, a case study model of a vascular closure device reported by Brandes *et al*[20] did not estimate QALYs and did not include a long-term model component. The authors state that this decision was taken because no data were available to estimate the impact of complications associated with the intervention and its comparators on health-related quality of life (HRQoL). Instead, a short-term decision tree was used to estimate an ICER

defined in terms of the additional cost of averted complications per catheterisation. Similarly, the early model for a new triple biomarker test for non-ST elevated myocardial infarction used the percentage of patients who are correctly discharged from the hospital as the main clinical endpoint for the model.[21] The decision to focus the analyses on these intermediate endpoints resulted in simpler model structures compared to a full model which estimates lifetime QALYs and costs with and without the intervention. None of the identified studies provided any additional recommendations on how to make other structural model simplifications. Restricting the outputs of an early model to intermediate outcomes would also limit the analyses which can be conducted using that model e.g., it would not be possible to conduct a full VOI analysis. It should also be noted that the intermediate outcome may not fully capture the full additional value (and potential adverse consequences) of the technology.

3.1.3.4 Identification of evidence to inform model parameters

Cosh *et al.*[15] and Hartz and John[22] state that choices regarding the selection of comparators included in an early model and the outcomes predicted by it (intermediate or final endpoints) are crucial. Cosh *et al.*[15] also state that the comparator in headroom analysis should always be the current gold standard method. These choices around the appropriate comparator(s) and whether long-term outcomes need to be modelled will influence what types of evidence will be needed to populate the model.

Regarding the data sources, the included papers refer to the use of evidence in early models obtained from published literature, early clinical evidence, small clinical studies, observational studies and expert opinion.[10, 11, 22] Of note, in their early model of the AtrialShaper device,[10] Pietzsch and Paté used animal testing data for the clinical effectiveness of the new device: for the length reduction of cardiac tissues exposed to radiofrequency electrodes, the researchers referenced the tissue shrinkage data from 3 studies(including 2 animal studies) and used weighted beta distributions to get an aggregated distribution for the achievable tissue shrinkage by AtrialShaper device. Hjelmgren *et al.*[17] suggest that using observational clinical data as the basis for modelling rather than data from randomised controlled trials (RCTs) could be less restrictive, more generalisable and may enable hypothetical comparisons between standard therapy and a variety of alternative strategies. Existing full models may also provide a source of evidence for parameters for new early models, either in terms of their individual parameter values (e.g. a health state utility value or the cost associated with treating a clinical event), or their predictions (e.g. lifetime QALYs or costs for a patient receiving a particular treatment).

Most of the papers included in the review acknowledge that owing to the early stage at which the economic analysis is conducted, evidence on the effectiveness or performance of the technology is likely to be limited or entirely absent and, therefore, the model outputs will be partly speculative and will frequently need to rely on expert opinion to quantify the uncertain parameters.[23] Based on early models developed for the NIHR Diagnostic Evidence Co-operatives (DECs), Abel and Shinkins[9] highlight that, for early diagnostic cost-effectiveness models, treatment effectiveness was a key determinant of cost-effectiveness in all cases, but that robust evidence on this aspect of the model was sparse.

3.1.3.5 Use of expert opinion

The included papers refer to three main uses of expert clinical input: (i) to inform the nature of the decision problem, e.g., defining the current and new care pathways, the positioning and design of the new technology and the target population; (ii) as a source for informing the model structure and assumptions (iii) as a source for the elicitation of estimates of uncertain parameter values, potentially including the effectiveness or performance of the technology where evidence is lacking.[11, 24]

Many of the included papers[5, 7, 24] make several suggestions around what to elicit and how values should be elicited., including structured expert elicitation. The precise methods suggested vary between the papers and, as they are not specific to early modelling, they are not discussed in detail here. However, of particular note, some authors suggested using elicitation to estimate specific model parameters in terms of point estimates plus ranges or distributions. Cao *et al.*[24] suggested combining elicitation with headroom analysis to obtain the probability distribution of the headroom available to a new device or technology. The approach described involves developing a baseline model of current care, eliciting probability distributions on the expected effect of the new intervention from clinical experts, then propagating uncertainty in all model inputs through the model in all model inputs through the model in order to derive the probability distribution of commercial headroom as a model output. This may be necessary where no evidence is available around the effectiveness or performance of the new technology.

Some authors[21, 25, 26] challenged the value of using expert elicitation to inform early economic models. Several concerns are raised including: (i) the absence of a universally agreed elicitation method; (ii) variability in results obtained using different approaches; (iii) difficulties in reaching consensus between experts and (iv) the substantial resource implications associated with conducting formal elicitation exercises which may not be feasible within the available timescales for model development.

3.1.3.6 Model evaluation methods

Owing to the high level of uncertainty around the effectiveness of the intervention, the need for extensive sensitivity analysis is recognised to be a key part of early economic evaluation.[9, 27, 28] Across the papers included in the review, various alternative analytic methods are proposed for use with early models, including headroom analysis and other types of DSAs, PSA and VOI analysis. With the exception of headroom analysis, these methods are commonly applied to the evaluation of full economic models.

Headroom analysis and other types of deterministic analyses

As noted in Section 3.1.3, many of the papers suggest the use of headroom analysis to explore the maximum price that could be achieved for a new technology given optimistic assumptions regarding the effectiveness of the new technology under consideration, or the necessary effectiveness of the technology given its anticipated cost. These analyses could include assumptions that the new technology has equivalent effectiveness to the current treatment in terms of mortality but offers an HRQoL benefit or they may be based on estimates of an effectiveness gap which describes the maximum potential increase in effectiveness or diagnostic accuracy of the new treatment compared with a current comparator. Buisman *et al.*,[11] suggests conducting threshold analyses to determine the minimum sensitivity and specificity at which a new medical test becomes an attractive alternative from an effectiveness standpoint: the different combinations of sensitivity and specificity estimates (ranging from 50% to 100%) are valued at a specified willingness-to-pay threshold. Abel and Shinkins *et al*.[9] suggest that given the high level of uncertainty around the effectiveness or performance of the new intervention, these types of analysis may be more meaningful to decision-makers than ICERs. Some authors have warned that the analysis of early models, in particular the assumptions underpinning headroom analysis, can be at risk of “pro-innovation bias” and “overoptimism bias” whereby the expected performance or effectiveness of the new technology are unrealistically inflated.[13]

Several authors also suggest the need for other types of deterministic sensitivity and scenario analyses in early models. These may include scenario analyses around the price of the new technology (which may not be relevant to the EVA process), or more broadly, the use of univariate DSAs around all model parameters.[11] Scholte *et al*.[3] has suggested that rather than presenting a base case analysis (as would usually be done in a full model evaluation), scenario analyses should instead be presented to explore the conditions under which the technology generates desirable cost-effectiveness outcomes.

PSA and VOI analysis

It has been argued that PSA is the most appropriate way to characterise parameter uncertainty, even for early models.[28] However, it has also been argued that DSAs might be considered to be more informative than PSA for a non-specialist audience in providing insight into factors affecting the cost-effectiveness of a new technology evaluated using an early model. It has also been suggested that undertaking PSA using early models could be misleading because many uncertainties in model parameters are hard to quantify and that this can lead to *"pseudo-certainty."*[4]

Several authors[7, 11, 22, 25, 29] suggest that VOI analysis should be carried out in the early stages of test development before a considerable amount of resource is committed to product development. However, the concerns described above regarding whether it is meaningful to conduct PSA using an early model will apply equally to VOI analysis.

Main points identified from the review

In summary, the papers included in the review suggests the following:

- There are many examples of early models. Whilst there is general agreement about what an early model is in terms of when the analysis is conducted, there is less clarity about how early models differ from full models.
- Simpler early models might be preferred because they are simpler and quicker to build than full models whilst still being considered "sufficient" for the interim decision they are designed to inform. Conversely, more complex early models akin to full models may be preferred as they can be re-used later when additional evidence has been collected.
- The papers highlight some ways in which an early model might be simpler than a full model, e.g., by limiting the model to intermediate outcomes, by excluding long-term outcomes and costs, or by assuming an artificially simplified pathway. Beyond these, the included papers did not offer guidance on how to build an early model or how to build simpler models.
- Many early models use the same types of modelling approaches as full models (decision trees, state transition models or combination of both).
- Existing models may be useful if they directly address the current decision problem, or if they can be used as a source of model parameters or part of the new model structure.
- Some authors suggest that elicitation of expert judgement may be useful, whilst others have challenged the feasibility and value of this approach. In the context of the constrained NICE EVA process, it is highly unlikely that formal elicitation techniques

will be either feasible or desirable. That is not to say that expert opinion of some type to inform model parameters will not be required, but that the methods to elicit and combine these views will necessarily be crude.

- There is general agreement that sensitivity analysis is required, but there is some disagreement about whether certain methods are useful - some authors seem to prefer PSA, whilst others prefer deterministic methods. Headroom analysis appears to be a key method for evaluating early models.

3.1.4 Pilot EVA review findings

This section provides a summary of the economic analyses conducted within all seven published EVAs as well as four EVAs which are in development. Table 1 summarises the main characteristics of each economic analysis, including the intervention, comparator and populations, the type of technology, whether a coded model was developed, the types of analyses conducted, and the economic outcomes generated by the model.

Table 1. Summary of pilot EVAs (published and in development)

NICE Topic ID	Interventions & Populations	Medical technology/ Diagnostic	Conceptual model (YES/NO)	Economic models (YES/NO) + Modelling approach	Types of economic analysis	Economic outcomes
HTE9	Digitally enabled therapies for adults with anxiety disorders *Compared against standard care*	Medical technology	YES	YES Decision tree	Cost-effectiveness analysis One-way sensitivity analysis Scenario analysis, VOI analysis	Net monetary benefit
HTE8	Digitally enabled therapies for adults with depression *Compared against standard care*	Medical technology	YES	YES Decision tree + Markov	Cost-utility analysis One-way sensitivity analysis Scenario analysis	Net monetary benefit
HTE7	Point-of-care tests for urinary tract infections to improve antimicrobial prescribing *Compared against urine dipstick + lab-based tests or lab-based testing alone*	Diagnostic	YES	YES. A simple model was developed but the EAG did not present any results.	N/a	N/a
HTE6	Genedrive MT-RNR1 ID kit to guide antibiotic use and prevent hearing loss in babies with bacterial infections *Compared against no testing*	Diagnostic	YES	YES Markov model using TreeAge Pro 2022	Cost-effectiveness analysis Deterministic sensitivity analysis	Costs per test kit Incremental costs per adverse event (AIHL) avoided Incremental costs per QALY gained
HTE5	ProKnow cloud-based system for radiotherapy data storage,	Medical technology	YES	NO	Cost-minimisation analysis Scenario analysis	Costs per technology per year Costs per patient per year

NICE Topic ID	Interventions & Populations	Medical technology/ Diagnostic	Conceptual model (YES/NO)	Economic models (YES/NO) + Modelling approach	Types of economic analysis	Economic outcomes
	communication and management *Compared against standard care*					
HTE4	Cari-Heart using AI to predict cardiac risk in patients with stable chest pain or suspected coronary artery disease *Compared against CTCA+ Clinical assessment*	Diagnostic	YES	NO	N/a	N/a
HTE3	Guided self-help digital cognitive therapy for children and young people with mild to moderate symptoms of anxiety or low mood *Compared against standard care*	Medical technology	YES	YES Decision tree using R	Cost-utility analysis One-way sensitivity analysis VOI analysis	Mean costs Mean QALYs Net monetary benefit
N/a	AI autocontouring to aid radiotherapy treatment planning *Compared against contouring methods in standard care*	Medical technology	NO	NO	Simple cost-consequence analysis	Resource use Time associated with the use of technology

NICE Topic ID	Interventions & Populations	Medical technology/ Diagnostic	Conceptual model (YES/NO)	Economic models (YES/NO) + Modelling approach	Types of economic analysis	Economic outcomes
N/a	AI software and chest X-ray for lung cancer *Compared against chest X-ray alone*	Diagnostic	YES	NO	Budget impact analysis	Anticipated budget impact for different populations (5 year)
N/a	Digitally enabled weight management programmes providing specialist weight management services *Compared against standard care*	Medical technology	YES	NO	Cost-utility analysis One-way sensitivity analysis Scenario analysis Threshold analysis	Costs QALYs Net monetary benefit
N/a	KardiaMobile_6 lead ECG Comparing with standard care	Diagnostic	NO *(EVA conducted by external assessment group [EAG]))*	NO	N/a	N/a
			YES *(Additional analysis conducted by NICE Decision Support Unit [DSU])*	YES Decision tree +Markov model	Cost-effectiveness analysis (using QALYs) Sensitivity analysis	Costs QALYs Incremental net monetary benefit

NICE - National Institute for Health and Care Excellence; HTE - health technology evaluation; N/a - not applicable; MT-RNR1 - mitochondrially encoded 12S ribosomal ribonucleic acid; AI - artificial intelligence; CTCA - CT coronary angiography; QALY - quality-adjusted life year; VOI - value of information; ECG - electrocardiogram; AIHL - aminoglycoside-induced hearing loss

Out of eleven EVA topics, six relate to medical technologies and five relate to diagnostic technologies. Amongst these EVAs, seven included standard practice as the comparator. Note that for KardiaMobile 6-lead ECG, the NICE DSU were asked to conduct an economic modelling exercise after the EAG had completed their review work. Therefore, there are twelve EVA reports for eleven topics including two versions of KardiaMobile's EVA.

A conceptual model was developed in ten EVA reports. Two reports (AI contouring to aid radiotherapy treatment planning assessment and the KardiaMobile_6 lead ECG assessment by EAC) did not develop conceptual models. Where a conceptual model was included these varied in the level of detail by which the proposed model was described. Five EVAs included the implementation of a coded economic model and reported results generated using that model. In one EVA (HTE7), the EAG developed a simple coded model in R but the type of analysis was not clearly reported and the results of these analyses were not presented in their assessment report. A decision tree was used in two medical technology EVAs, a Markov model was used in one diagnostic EVA, a hybrid decision tree and Markov model was used in one medical technology and one diagnostic EVA. All EVAs which included a fully coded model were cost-utility analyses and reported health effects in terms of QALYs. All economic models reported based case analyses, DSAs and scenario analyses. VOI analysis was conducted using two medical technology models (HTE3 and HTE9). Headroom analysis was not reported in any EVA. Only one EVA report included a threshold analysis.

In developing economic models, five of EVAs referred to the use of previous models or economic evaluations to inform the model structure, downstream events, outcomes and costs of interventions. In the early assessment of KardiaMobile by NICE DSU,[30] two different previous models are incorporated into the early model: one for decision tree component and one for state-transition model component. Some EVAs used the expert input to inform clinical care pathways, downstream clinical events and costs (e.g., HTE5, HTE6, HTE7 and AI derived software for chest X-rays EVA). None of the EVA reports described the use of formal elicitation methods of expert judgement to obtain estimates of parameter values.

Seven EVAs did not include a coded full model (HTE7 had a partially coded model). Among them, one medical technology EVA (HTE5) instead presented a cost-minimisation analysis, one medical technology EVA presented a simple cost-consequences analysis and one diagnostic EVA presented a budget impact analysis. Amongst those EVAs which did not include a full model, suggested approaches for future models included decision trees, hybrid decision tree and Markov models and discrete event simulation (DES). Across these EVAs, reasons given for not developing a coded economic model included:

- Limited data on the clinical effectiveness of the new intervention
- Limited data on consequences of interventions (e.g., lack of data on impact of antibiotics prescription as a result of different diagnostic tests for urinary tract infections [UTIs])
- The absence of previous models or economic evaluation studies comparing the new intervention and the comparator, and the lack of evidence on costs and cost-effectiveness of the new intervention
- Previous trials of the same technology assessing different clinical outcomes in different populations of interest/ previous models looking at a different comparator in other populations,
- Clinical experts' suggestions (e.g., in ProKnow EVA, there are inconsistencies in terms of treatment planning between different local departments, and early modelling by taking into account of those variations may overcomplicate the pathway and assumptions would not be appropriate).

Across the full range of EVAs, the following key areas of uncertainty were identified:

- Resource use (e.g., licence costs, healthcare professional costs based on the grade and the duration of the procedure)
- Clinical effectiveness including both intermediate and long-term effects
- Length of treatment courses, duration of interventions and level of support from the health workers, particularly for medical technologies EVA (e.g., in digitally enabled mental health interventions)
- Diagnostic accuracy of tests (e.g., real-world sensitivity of Genedrive RNR1 ID because of the uncommon nature of the disease, to include failure rates as in AI-derived software)
- Prevalence of the predisposing conditions and disease/clinical events in the population of interest, especially in the case of diagnostic tests for rare diseases/conditions (e.g., the Genedrive MT-RNR1 ID diagnostic test)
- Clear definition or agreement on standard practice (e.g., the ProKnow radiotherapy EVA)
- Clinical outcomes of diagnostic tests-derived treatment changes (e.g., the impact of antibiotics prescribing in the POCT EVA and the impact of statins dose changes in Cari-Heart EVA)

In the NICE guidance reports of the three EVAs, it was mentioned explicitly that the committee considered their economic analyses or exploratory modelling to make the final recommendation about the implementation of the technologies in practice. For example, in HTE6, the committee concluded that based on early economic model

results, the Genedrive test had the potential to be cost-effective over a lifetime, in HTE7, the committee concluded that there were uncertainties in the exploratory modelling of EAG regarding estimation of the test-derived treatment changes to recommend early routine use of the intervention in the NHS, and in the HTE9, the committee considered that despite some limitations in the model, the new interventions are less likely to cost less than standard care due to less therapist time than the standard care.

4 DISCUSSION AND RECOMMENDATIONS FOR FUTURE EVAS

Our review of existing literature shows that many applications of “early modelling”, and guidance on potential methods, relate to different settings to those faced by NICE and the External Assessment Groups submitting evidence. “Early modelling” is often undertaken during the technology development stage to inform evidence collection efforts and potential pricing decisions. It is more often linked to this developing and amendable series of questions and, as such, may be thought of as part of an ongoing process. The NICE EVA setting is somewhat different in that the focus here is on the alignment of early modelling to inform a single committee decision, albeit one that may be revisited at some point in the future. This means that there may be more information about the indication being targeted though not the precise role in the care pathway, and a price, or likely price range for the technology, is known. This occurs at a time when it is known a priori that the core clinical evidence is insufficient for making a definitive decision about recommending for routine adoption, and that there are significant limitations on the time and resources available to undertake the analysis. These factors raise specific challenges for those tasked with submitting cost effectiveness evidence to NICE as part of the EVA process.

Time and resource constraints mean that choices must be made regarding where to focus efforts around the cost effectiveness assessment. It is unlikely to be a reasonable expectation for each of the activities that would be considered standard in a full cost effectiveness assessment to be conducted, or conducted to the same degree of rigour, except in those rare cases where the model and associated parameters are extremely simple. The prior knowledge that there is little or no relevant evidence relating to key aspects of clinical effectiveness raises additional considerations about how to conduct economic evaluation and the information such analyses can convey for decision makers.

The aims of economic modelling for NICE EVA may be:

a) To make an assessment of whether it is plausible that the technology in question may be cost-effective.

b) To identify the key areas of uncertainty where resolution or reduction of that uncertainty would be of most value to a future assessment and the potential consequences of an incorrect adoption decision.

The weight given to answering these different aims will, in part, inform the choices around the economic analysis. And, in turn, the nature of the technology, the key sources of uncertainty and the availability of other relevant information for the economic model will inform the degree to which each of these aims can be addressed within the short timescales and resources available. For example, in the case of diagnostic technologies, if there is both an absence of evidence on the diagnostic performance of the technology in question, and a complex and unquantified set of downstream costs and health impacts that arise from diagnostic information, then it will be much more difficult to address aim a) with any meaningful degree of confidence. Similarly, for medical technologies with a therapeutic purpose where there is a complex or unclear link between outcomes where evidence is likely to exist and the health and resource use impacts of core interest, economic modelling that informs a) will be challenging.

It should also be noted that the range of decision making options available to the committee are limited to conditional approval with additional research ("Use while further evidence is generated"), recommended only in the context of research, or not recommended for use.

A flowchart of the options available to Assessment Groups undertaking EVAs is provided in Figure 1. The following text expands on these different options and outlines at least some of the issues that are important to consider at each stage. But there is no rigid form in which these issues can be considered. Contextual factors, specific to each case, will impact the decisions analysts make and therefore the approach to modelling that is decided upon. The guidance here is intended solely to assist Assessment Groups as they make their own judgements about the most appropriate approach to take.

We illustrate our discussion using diagnostic technologies but note that many of these same issues apply in the case of tests more generally (for example, those situations where the technology has a prognostic or predictive focus) and to those technologies that have a therapeutic purpose. This is particularly the case where the care pathway that leads to cost or health benefits is complex and needs to be modelled from some intermediate outcome. In the case of diagnostic technologies, it is typical to consider three broad components of a cost effectiveness model:

i) the impact of the technology in classifying patients conditional on their true status. Typically, this can be thought of in terms of sensitivity and specificity of the test and the underlying prevalence of the condition in the population of interest, though this might also include technologies which assign patients to risk categories.

ii) the impact of that classification on the management and treatment decisions taken.

iii) the downstream costs and health benefits that accrue based on those management decisions.

A pragmatic review of the literature seeking existing models, sources of parameter values and evidence about the current care pathway should be performed (Labelled 1 in the flowchart). This should consider clinical guidance, including that issued by NICE, pertaining to downstream interventions. On rare occasions, this will identify an existing, relevant cost effectiveness model that is considered suitable for all three elements (2). This might be the case if a competitor diagnostic technology has previously been assessed, or if an economic evaluation is available in the public domain for other reasons. In the ideal situation, the functioning decision model can be obtained and is considered to be of a sufficient quality for informing the current decision. In this case, there may only be minimal adaptations required to address the issue of potential cost effectiveness of the new technology. This would require the inclusion of appropriate arms for the technology and comparators but, in this ideal case, the focus of the adaptation would be on the costs specific to the technology, any adverse events from the technology itself (aside from its diagnostic accuracy), and specific information on the diagnostic accuracy. Given that this latter issue is likely to be one area of substantial uncertainty, the model could be evaluated using plausible ranges for these diagnostic parameters in sensitivity analyses. Analytical time may include the generation of evidence to inform these plausible ranges, for example by detailed review of relevant existing literature (perhaps from similar technologies, in other populations) and/or via expert opinion. However, even where such short-cut modelling options exist, the degree of uncertainty in the evidence relating to stage ii) and iii) may still be considered substantial. There may be good reasons why the impact on management decisions of a new technology is not straightforward to assess and cannot be inferred reliably from existing approaches or their evidence on sensitivity and specificity (for example, an inherent cautiousness amongst clinicians interpreting diagnostic information from a new technology type. Or those situations where diagnostic information from the technology is combined with other clinical factors to determine management recommendations). A related issue here is the need to reflect the consequences of misclassification and, in particular, the length of time that may elapse before these errors are identified and the remedial action that is taken once they are.

This “ideal” situation rarely exists. In those situations where a relevant model exists, the executable version of the model may not be accessible. Open-source models may diminish this occurrence in the longer term. In principle, publications should provide information that is sufficiently detailed to allow replication. In practice, it is often not feasible to reproduce all aspects of a model and replicate its results accurately, particularly within a short timeframe.

Available models may require a significant degree of adaptation because they do not reflect relevant aspects of current practice or current understanding of the disease, are based on historical evidence or were designed for decision makers adopting a different perspective or jurisdiction.

In what is likely to be the most frequently occurring situation, where no model of the diagnostic pathway exists (covering components i - iii above), there is a requirement for a conceptual model of the diagnostic and management approach to be built as a first stage to estimate cost effectiveness (3).

Components i and ii are often relatively simplified representations of the pathway in diagnostics models, constructed using decision trees. However, there can be a significant investment of analyst time required to fully understand the pathway, relevant testing strategies (such as repeat or combinations of tests) and the potential impact on clinical decision making. In the absence of clinical studies that set out a diagnostic pathway and test the impact of the technology and its comparators on these aspects of management, considerable reliance on clinical experts will be required. Of course, even where such studies do exist, it may be the case that they do not align with the way in which the technology may work in NHS practice. Clinical input is always required. Judgements will be needed to ascertain the types of trade-offs between simplification of the pathway, focussing on specific aspects of the decision problem and the complexities of clinical pathway. For example, in the case of “6 Lead ECG for measuring cardiac QT interval in people having antipsychotic medication”, simplifying decisions were made to concentrate on the comparison of single use of 6 lead vs the 12 lead comparator, rather than strategies that modelled repeat use.

Component iii, the set of downstream costs and health benefits that accrue as a function of management decisions, can be complex and time consuming to model robustly. In the EVA setting, this is the situation where short-cut options are likely to be required to answer the required questions in the cost effectiveness analysis. Several short-cuts may be possible. First, there may be an existing model of these downstream costs and benefits that can be adapted for use (4). Typically, this would be feasible if there has been assessment and guidance issued for the treatments that would be used following diagnosis. Where such a

model can be obtained or replicated, this is likely to provide estimates of most relevance for true-positives. Simple adaptations (if any are needed at all) can be considered to quantify the costs and health benefits such as for those treated that do not have the condition (false-positives) and those with the condition not receiving treatment for a period (false-negatives). In many situations, it may be reasonable to assume that false negatives will be diagnosed correctly shortly after the initial test, due to ongoing symptoms and the use of other investigations, with no health impacts and easily identified cost implications. In those situations where this delay leads to adverse health impacts, relevant cost and outcome estimates may be available for patient subgroups with more advanced disease.

The degree of adaptation required will depend on the specific case. For example, in the KardiaMobile assessment, researchers developed a state-transition model as the second component of their decision model by replicating an existing treatment model from a NICE clinical guideline and adapted it to address some additional outcomes (treatment-related adverse events and disease mortality) and costs. It may also be possible to replicate certain portions of an existing model to generate the information required for the diagnostic model (e.g., reproducing reported survival functions to estimate mean survival time or per-cycle mortality risk in one treatment group, without estimating other outcomes or costs).

It may be possible to avoid the need for replication of an executable model here but to instead simply use the reported outcomes from a modelling exercise (i.e., the mean costs and mean QALYs) as payoffs in the diagnostic decision tree (5). When taking this approach there is a need to ensure that the intended model populations overlap sufficiently in relevant areas for results to be considered generalisable and that the analytical methods are consistent with those used by NICE (for example in relation to perspective and discounting). Where relevant models have been produced as part of the production of NICE guidance or clinical guidelines, this may provide reassurance relating to both the methods that have been used and the quality assurance steps that the modelling work will have been subjected to. For models obtained from elsewhere, including from peer reviewed publications, this may be less clear. Some degree of validation of the outputs can be undertaken where outputs are reported in a way that allows such an assessment to be made (for example by reporting numbers of patients in health states, or the frequency with which specific events are estimated to occur. These rates can be compared to UK NHS experience reported in other studies or by seeking clinical judgement). In the conceptual modelling of an EVA pilot for the diagnostic technology using AI software to analyse Chest X-rays in lung cancer, the above approach was mentioned by researchers: they recommended assigning the long-term treatment costs and utility values over a specific time horizon at the end of the decision tree in their decision model. As above, it is likely that these reported summary outcomes will be insufficient for the needs of the

diagnostic model and adaptations will be required. For example, it may be considered that false-positive cases are assigned to treatment, at least for a period of time, and be subject to the risks of adverse events without the benefits of treatment efficacy. Where the frequency, impact on mortality risk and/or health-related quality of life and / or costs for adverse events are reported, these adjustments may be relatively simple to perform. Exactly how such adjustments can be made will depend on the specific case and the information that is contained in summary model outputs to make them.

In many situations, no relevant existing model or model outputs will be identified. Analysts in this situation will need to determine if a suitable downstream model can be developed and populated in the available timeframe (6). This may be feasible if a simplified pathway is considered appropriate and / or if suitable parameter values are easily available. In a full evaluation, the absence of relevant evidence may prompt the use of expert opinion. Formal methods for eliciting parameters and associated uncertainty from sufficiently informed samples of experts have been documented but these are themselves time consuming exercises. Expert opinion may be an option in these EVA situations, but it is likely to be elicited in a much less formal manner at this stage because of the short period available to conduct the analysis. For example, very small numbers of relevant experts may be consulted and asked to provide simple point estimates for relevant parameters, with limited characterisation of uncertainty. The risks of introducing bias from this simplified approach needs to be acknowledged.

It is often valuable to attempt to develop simplified models of these downstream effects. If we cannot reliably distinguish differences in costs and outcomes between the different diagnostic outcomes, then we cannot assign value to a more or less specific/sensitive test. In addition to uncertainty around diagnostic accuracy, such modelling exercises can help to identify the importance of factors that would not have been envisaged without quantification of the decision problem using a model. For example, the importance of adverse events and particularly cardiac risk associated with different anti-psychotic drugs is of crucial importance in the 6 lead ECG example cited above. However, with excessive simplification of a model structure and the non-systematic identification of parameter values, there is also a risk of drawing erroneous conclusions both of the potential for a technology to be cost-effective, but also the key drivers of value. Therefore, in this situation, there is a clear judgement to be made by analysts. They must assess the trade-offs between greater simplification of the model structure in order to produce a set of estimates, versus maintaining a more detailed, accurate, description of the downstream care pathway in a conceptual model that does not produce outputs. The latter can be informative when used to highlight the likely key drivers of value, contrasted with a summary of where there appear to be significant gaps in the existing

evidence. The approach can inform judgements about candidate issues for future research in a way that highly simplified downstream models, or “bolting-on” costs, QALYs or other summary measures of outcome from external models, cannot.

Where it is judged that these downstream costs and health effects cannot be modelled to the required degree of rigour, the reasons for this should be documented, and consideration given to the insights that can be gained from a model of the diagnostic phase only (stage i above) based on prevalence and diagnostic accuracy information, with a conceptual model developed to reflect the downstream impacts (stages ii and iii) (7). This may occur in those situations where the impact of test results on management of patients is unclear, or the care pathway is complex and non-standardised, for example. In many situations, the distinction may be between a simple model of stages i and ii, with the downstream impacts (stage iii) lacking. In this situation, models may reflect the additional direct costs from the use of the new technology and the impact on intermediate outcomes such as cases detected. This information alone is likely to be much less valuable for decision makers than a model of all stages but will provide supplementary insights to decision makers than those obtained from information on diagnostic accuracy alone.

It will often be the case that there is a clear link between test results and expected changes in the management of patients (stage ii), or that clinical opinion can be obtained to estimate these impacts. In this situation, model outputs would provide estimates of the proportions of patients allocated to each management option conditional on their true status. This contrasts with the proportions of patients by test outcome and conditional on true status.

Coupled with a narrative account of the key sources of downstream costs and health benefits, with indicative values for those categories that can easily be identified, this will give some useful information relating to aims a and b above. Certain threshold type analyses could be performed. For example, if the cost of the new technology is known and it is possible to generate estimates of additional or offset costs associated with treatment, how many additional QALYs would be required to achieve a positive net health benefit? (This is linked to the idea of “headroom analysis” referred to in existing literature on early modelling but focusses on effectiveness rather than cost). This would be feasible, or at least provide more reliable estimates, in those situations where net costs are considered to be impacted significantly only by a small number of categories where changes in management decisions can be simply estimated. In those situations where it is not considered feasible to estimate downstream costs, the model of stages i and ii could be used to quantify the net health benefits accrued in the short term and therefore the magnitude of the net health benefits required downstream to offset this.

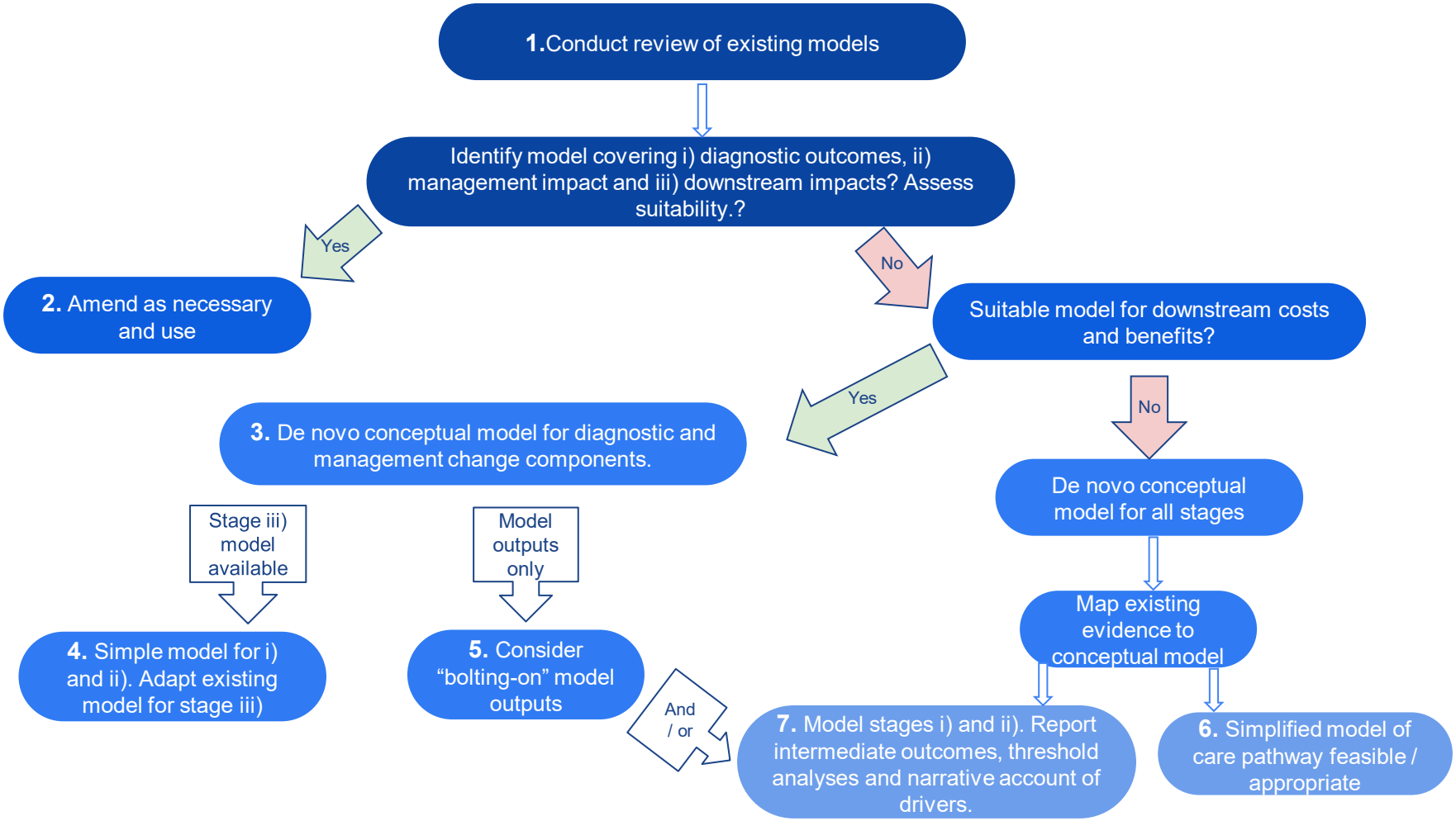
The steps outlined here seek to provide advice that can maximise the usefulness of modelling within the resource, time and evidence constraints of the EVA process. The potential steps that can be used to simplify aspects of the modelling can guide but not substitute the judgement of experienced analysts and will need to be adapted on a case by case basis. These steps have been highlighted by reference to diagnostic tests but those same considerations remain relevant for other technology types, such as medical technologies with a therapeutic value, or diagnostics where value is generated from outcomes such as the time to diagnosis and not improved diagnostic accuracy. In the absence of an existing model, the development of a conceptual model will always be the natural starting point in these scenarios. The distinction between elements i, ii and iii in the diagnostic pathway may be less relevant in these settings but similar challenges may be encountered for alternative parts of the pathway to be modelled.

In all these modelling scenarios, it is inevitable that parameter estimates are based on non-systematic, comprehensive summaries of available evidence and expert opinion that itself may be subject to bias. This increases the probability that recommendations are misguided, leading to potentially misplaced interim adoption decisions and the implementation of costly research recommendations. The limitations and risks need to be clearly outlined and, where feasible, the available model framework used to indicate the implications of decision error.

Summary of key recommendations

I. Conduct a pragmatic review of available models
II. If a suitable model covering all stages can be identified:
    a. Obtain executable model if feasible
    b. Consider generalisability of methods and patient population to current decision problem
    c. Use or replicate with suitable modifications
III. If a suitable model covering all stages cannot be identified, then:
    a. If a suitable model does exist for the treatment consequences component of the model then develop a de novo conceptual model of the diagnostic and management stages
    b. If no suitable model exists then develop a de novo conceptual model for all stages of the decision problem.
    c. Conceptual models should seek to describe the full relevant aspects of the care pathway, and parameters, that would be needed in a full evaluation. This helps identify likely key value drivers and where estimates from existing data may be lacking.
IV. Following on from 3a), develop an executable model focussing on simplified diagnostic pathways if necessary and adapting the available existing model of downstream costs and QALYs
V. Following on from 3b), develop a simplified version of the conceptual model highlighting where time and resource constraints impose the need for compromises such as
    a. Limiting consideration to a small number of routes or highly simplified characterisations of the care pathway
    b. Focussing on the diagnostic aspect of the model and either “bolting -on” estimates of downstream costs and benefits from external estimates, or reporting interim outcomes
VI. Expert opinion can be elicited in informal ways to parameterise the model. Uncertainty and potential biases should be reflected and discussed.
VII. Standard approaches to reflect uncertainty, particularly deterministic sensitivity analysis based on ranges considered plausible or extreme value testing, should be reported. Threshold analyses are likely to be particularly informative.
VIII. Consider the key limitations in any executable model and provide information on the potential likelihood and implications of guidance, including relating to research priorities, that is subsequently shown to be incorrect.

Figure 1. Summary flowchart of EVA modelling options



REFERENCES

1. National Institue for Health and Care Excellence. Early value assessment interim statement London: NICE 2022.
2. National Institute for Health and Care Excellence. Early Value Assessment (EVA) for medtech. NICE; 2023. Available from https://www.nice.org.uk/about/what-we-do/eva-for-medtech (Accessed 31/08/2023).
3. Scholte M, Marchau V, Kwakkel JH, Klijn CJM, Rovers MM, Grutters JPC. Dealing With Uncertainty in Early Health Technology Assessment: An Exploration of Methods for Decision Making Under Deep Uncertainty. *Value Health* 2023;26:694-703.
4. Grutters JPC, Govers T, Nijboer J, Tummers M, van der Wilt GJ, Rovers MM. Problems and Promises of Health Technologies: The Role of Early Health Economic Modeling. *Int J Health Policy Manag* 2019;8:575-82.
5. IJzerman. MJ, Koffijberg H, Fenwick E, Krahn M. Emerging Use of Early Health Technology Assessment in Medical Product Development: A Scoping Review of the Literature. *Pharmacoeconomics* 2017;35:727-40.
6. Vallejo-Torres L, Steuten LM, Buxton MJ, Girling AJ, Lilford RJ, Young T. Integrating health economics modeling in the product development cycle of medical devices: a Bayesian approach. *Int J Technol Assess Health Care* 2008;24:459-64.
7. Love-Koh J. How Useful Are Early Economic Models? Comment on "Problems and Promises of Health Technologies: The Role of Early Health Economic Modelling". *Int J Health Policy Manag* 2020;9:215-7.
8. Markiewicz K, van Til JA, MJ IJ. Medical devices early assessment methods: systematic literature review. *Int J Technol Assess Health Care* 2014;30:137-46.
9. Abel L, Shinkins B, Smith A, Sutton AJ, Sagoo GS, Uchegbu I*, et al.* Early Economic Evaluation of Diagnostic Technologies: Experiences of the NIHR Diagnostic Evidence Co-operatives. *Med Decis Making* 2019;39:857-66.
10. Pietzsch JB, Paté-Cornell ME. Early technology assessment of new medical devices. *Int J Technol Assess Health Care* 2008;24:36-44.
11. Buisman LR, Rutten-van Mölken MP, Postmus D, Luime JJ, Uyl-de Groot CA, Redekop WK. THE EARLY BIRD CATCHES THE WORM: EARLY COST-EFFECTIVENESS ANALYSIS OF NEW MEDICAL TESTS. *Int J Technol Assess Health Care* 2016;32:46-53.
12. Grabowski H. The Effect of Pharmacoeconomics on Company Research and Development Decisions. *PharmacoEconomics* 1997;11:389-97.
13. Fasterholdt I, Krahn M, Kidholm K, Yderstræde KB, Pedersen KM. Review of early assessment models of innovative medical technologies. *Health Policy* 2017;121:870-9.
14. IJzerman. MJ, Steuten LM. Early assessment of medical technologies to inform product development and market access: a review of methods and applications. *Appl Health Econ Health Policy* 2011;9:331-47.
15. Cosh E, Girling A, Lilford R, McAteer H, Young T. Investing in new medical technologies: A decision framework. *Journal of Commercial Biotechnology* 2007;13:263-71.
16. E.A.Boudewijns, Otten TM, Gobianidze M, Ramaekers BL, van Schayck OCP, Joore MA. Headroom Analysis for Early Economic Evaluation: A Systematic Review. *Appl Health Econ Health Policy* 2023;21:195-204.
17. Hjelmgren J, Ghatnekar O, Reimer J, Grabowski M, Lindvall O, Persson U*, et al.* Estimating the value of novel interventions for Parkinson's disease: an early decision-making model with application to dopamine cell replacement. *Parkinsonism Relat Disord* 2006;12:443-52.
18. Retèl VP, Grutters JPC, van Harten WH, Joore MA. Value of research and value of development in early assessments of new medical technologies. *Value in health : the*

journal of the International Society for Pharmacoeconomics and Outcomes Research 2013;16:720-8.
19. Kolominsky-Rabas PL, Kriza C, Djanatliev A, Meier F, Uffenorde S, Radeleff J*, et al.* Health Economic Impact of a Pulmonary Artery Pressure Sensor for Heart Failure Telemonitoring: A Dynamic Simulation. *Telemed J E Health* 2016;22:798-808.
20. Brandes A, Sinner MF, Kääb S, Rogowski WH. Early decision-analytic modeling - a case study on vascular closure devices. *BMC Health Serv Res* 2015;15:486.
21. Kip MM, Steuten LM, Koffijberg H, MJ IJ, Kusters R. Using expert elicitation to estimate the potential impact of improved diagnostic performance of laboratory tests: a case study on rapid discharge of suspected non-ST elevation myocardial infarction patients. *J Eval Clin Pract* 2018;24:31-41.
22. Hartz S, John J. Contribution of economic evaluation to decision making in early phases of product development: a methodological and empirical review. *Int J Technol Assess Health Care* 2008;24:465-72.
23. Girling A, Young T, Brown C, Lilford R. Early-Stage Valuation of Medical Devices: The Role of Developmental Uncertainty. *Value in Health* 2010;13:585-91.
24. Cao Q, Postmus D, Hillege HL, Buskens E. Probability elicitation to inform early health economic evaluations of new medical technologies: a case study in heart failure disease management. *Value Health* 2013;16:529-35.
25. Steuten LM, Ramsey SD. Improving early cycle economic evaluation of diagnostic technologies. *Expert Rev Pharmacoecon Outcomes Res* 2014;14:491-8.
26. Bojke L, Soares M, Claxton K, Colson A, Fox A, Jackson C*, et al.* Developing a reference protocol for structured expert elicitation in health-care decision-making: a mixed-methods study. 2021;25:37.
27. Sutton AJ, Lamont JV, Evans RM, Williamson K, O'Rourke D, Duggan B*, et al.* An early analysis of the cost-effectiveness of a diagnostic classifier for risk stratification of haematuria patients (DCRSHP) compared to flexible cystoscopy in the diagnosis of bladder cancer. *PLOS ONE* 2018;13:e0202796.
28. Federici C, Torbica A. Expanding the Role of Early Health Economic Modelling in Evaluation of Health Technologies Comment on "Problems and Promises of Health Technologies: The Role of Early Health Economic Modeling". *Int J Health Policy Manag* 2021;10:102-5.
29. Hall PS, Smith A, Hulme C, Vargas-Palacios A, Makris A, Hughes-Davies L*, et al.* Value of Information Analysis of Multiparameter Tests for Chemotherapy in Early Breast Cancer: The OPTIMA Prelim Trial. *Value Health* 2017;20:1311-8.
30. Metry A, Gillet M, Wailoo A. KardiaMobile 6L for measuring QT interval in people having antipsychotic medication.Early Value Asssessment Cost Effectiveness Analysis. NICE DSU Report.; 2023.