Early economic evaluation of diagnostic technologies: experiences of the NIHR Diagnostic Evidence Co-operatives

Lucy Abel[1*], Bethany Shinkins[2,3], Alison Smith[2,3], Andrew J Sutton[2,3], Gurdeep S Sagoo[2,3], Ijeoma Uchegbu[4], A. Joy Allen[5], Sara Graziadio[5,6], Eoin Moloney[7], Yaling Yang[1◦], Peter Hall[2,3,8◦]

[1] Nuffield Department of Primary Care Health Sciences, University of Oxford

[2] The NIHR Diagnostic Evidence Co-operative Leeds, Leeds Teaching Hospitals NHS Trust

[3] Test Evaluation Group, Academic Unit of Health Economics, University of Leeds

[4] The NIHR Diagnostic Evidence Co-operative London, Imperial College London

[5] The NIHR Diagnostic Evidence Co-operative Newcastle, Newcastle University

[6] Newcastle upon Tyne Hospitals Foundation Trust

[7] Health Economics Group, Institute of Health and Society, Newcastle University

[8] Cancer Research UK Edinburgh Centre, University of Edinburgh

*Corresponding author

◦These authors contributed equally

An early draft of this paper was presented at the Winter 2018 Health Economics Study Group, City University, UK.

Word count: 5,196

Running Title: Early economic evaluation of diagnostics

Disclosure

Financial support for this study was provided in part by a grant from the National Institute for Health Research (NIHR). The funding agreement ensured the authors' independence in designing the study, interpreting the data, writing, and publishing the report.

This project was funded by the NIHR Community Healthcare MedTech and In-Vitro Diagnostics Co-operative (MIC) Oxford, the NIHR Leeds MIC, the NIHR London MIC, and the NIHR Newcastle In-Vitro Diagnostics Co-operative, and supported by the NIHR Biomedical Research Centre, Oxford. The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.

A number of projects on which this paper is based were funded by the Small Business Research Initiative (SBRI).

ABSTRACT

Diagnostic tests are expensive and time consuming to develop. Early economic evaluation using decision modelling can reduce commercial risk by providing early evidence on cost-effectiveness. The National Institute for Health Research Diagnostic Evidence Co-operatives (DECs) were established to catalyse evidence generation for diagnostic tests by collaborating with commercial developers; DEC researchers have consequently made extensive use of early modelling.

The aim of this paper is to summarise the experiences of the DECs using early modelling for diagnostics. We draw on eight case studies to illustrate the methods, highlight methodological strengths and weaknesses particular to diagnostics, and provide advice.

The case studies covered diagnosis, screening and treatment stratification. Treatment effectiveness was a crucial determinant of cost-effectiveness in all cases, but robust evidence to inform this parameter was sparse. This risked limiting the usability of the results, although characterisation of this uncertainty in turn highlighted the value of further evidence generation. Researchers evaluating early models must be aware of the importance of treatment effect evidence when reviewing diagnostics cost-effectiveness.

Researchers planning to develop an early model of a test should also: 1) consult widely with clinicians to ensure the model reflects real-world patient care, 2) develop comprehensive models that can be updated as the technology develops, rather than taking a “quick and dirty” approach that may risk producing misleading results, and 3) use flexible methods of reviewing evidence and evaluating model results, to fit the needs of multiple decision makers.

Decision models can provide vital information for developers at an early stage, although limited evidence mean researchers should proceed with caution.

WORD COUNT = 259

INTRODUCTION

Medical tests are ubiquitous in modern medicine. They can be used not only to diagnose disease, but also to monitor patients and provide a prognosis. With the rise of screening programmes, decision tools and guidelines, medicine is coming to rely evermore on accurate medical tests (1).

New tests require evaluation to assess cost-effectiveness, but this is a more complex proposition than for new treatments, which have a direct effect on patient outcomes. Medical tests instead tend to affect patient health only indirectly, by guiding the choice of treatment used (1).

The cost-effectiveness of a test therefore depends not only on diagnostic accuracy, but also the downstream impact of a given diagnosis on patient management, and in turn the cost-effectiveness of interventions used. At the same time, many tests are developed by small biotechnology companies who lack the resources to conduct trials with a relevant primary health outcome (2). In this high-risk commercial landscape, test developers may benefit from using decision modelling methods to assess the potential for cost-effectiveness earlier in the research and development (R&D) pipeline.

Early economic evaluation provides an initial assessment of whether a technology has the potential to be cost-effective, and under which conditions, before significant resources have been invested in its development. An early economic model typically relies on the literature, expert opinion, and early-phase test-specific clinical evidence to build a decision model that represents the intended clinical pathway for the technology under evaluation. Many of the model parameters are likely to be subject to great uncertainty, but scrutiny of the model can identify which model parameters are important determinants of final cost-effectiveness (3).

Early modelling is not intended to guide the adoption decisions made by organizations such as the National Institute for Health and Care Excellence (NICE), but instead to inform those engaged in the research and development process, such as manufacturers, investors and public funding bodies. On the basis of early cost-effectiveness evidence, a developer has the opportunity to refine their technology to better meet clinical need, reposition their technology within the same or an alternate care pathway, or abandon a diagnostic technology which is highly unlikely to be cost-effective (4). It is thought that early modelling has been used extensively in pharmaceutical R&D, although as most of these models are never published, exact numbers are unknown (3).

The National Institute for Health Research (NIHR) Diagnostic Evidence Co-operatives (DECs)

The NIHR DECs were established in 2013 (superseded by the NIHR MedTech and In vitro diagnostic Co-operatives (MICs) in 2018) and provided £4 million funding over four years across four institutions (Imperial College London, University of Leeds, University of Oxford and Newcastle University), in partnership with local NHS organisations. Their aim was to make clinical and academic expertise available to developers of medical tests, in order to catalyse the generation of high-quality evidence (5). A substantial proportion of this work involved collaborating with technology developers as independent partners on evidence generation projects. The resulting eight projects, all of which involved early economic modelling, form the basis for this discussion paper.

Table 1 summarises the decision problems considered. Most tests were being developed for secondary care settings, with cancer being the most common clinical area (4 cases), followed by infection (two cases). Five of the tests were in-vitro diagnostics evaluating biomarkers from patient

samples, two were genetic tests, with one imaging technique. Four of the evaluations have been published.

The aim of this paper is to describe the experiences of the NIHR DECs in applying early economic modelling to diagnostics and discuss its strengths and challenges. The structure of the paper is as follows: the first section outlines the modelling projects in question. The following section describes our experiences using early modelling methods, and what we consider to be the strengths and challenges of these methods. We then explore the outstanding questions around the methodological soundness of early modelling for diagnostics. Finally, we offer guidance on how to maximize the usefulness of these methods, before concluding.

Table 1: Summary of NIHR DEC-Industry collaborative early economic modelling projects

DEC location	Population	Setting	Index test	Aim of test	Current status of test	Comparators	Citation
Leeds	Patients at high risk of colorectal cancer referred for colonoscopy as a result of a positive faecal occult blood test	Secondary care	EMI-137 imaging agent	More accurate detection of/ better visualisation of small polyps and flat lesions	Under development	Standard white Light (WL) colonoscopy	
Leeds	Patients at high risk of urothelial bladder cancer presenting to haematuria clinics	Secondary care	Diagnostic classifier for risk stratification of haematuria patients (DCRSHP) biochip	Triage diagnostic test for risk stratification of haematuria patients	Under development	Flexible cystoscopy (Randox, 2016)	(7)
Leeds	Patients with newly diagnosed metastatic colorectal cancer	Secondary care	KRAS Oncobeam (Sysmex)	Identify somatic KRAS mutation to guide treatment with EGFR antibody	Under development, entering validation	Pyrosequencing	
Leeds	Patients admitted to critical care, at risk of Acute Kidney Injury (AKI)	Secondary care	Multiple, including NGAL, Cystatin-C and Nephrockeck ®	Early diagnosis or risk stratification of AKI	Various	Standard care diagnosis based on serum creatinine and urine output	(8)
Leeds	Patients surgically treated for early breast cancer	Secondary care	Oncotype DX (Genomic Health Ltd)	Predict benefit from adjuvant chemotherapy	NICE approved	NHS PREDICT, Prosigna®, IHC4, MammaPrint	(9)

London (Imperial college)	Patients with blood clotting issues, particularly those with major haemorrhage (excessive bleeding) presenting within the trauma unit	Secondary care	CoaguScan	Global coagulation assay - measures the ability of a patient's blood to clot (i.e. their current clotting status)	Under development: Clinical validation study among healthy population, in trauma and in the pre-hospital environment	TEG, ROTEM though not enough evidence, according to NICE, for application in trauma.	
Newcastle	Patients at high risk for Carbapenemase-producing Enterobacteriaceae (CPE) upon admission to UK hospitals	Secondary care	PCR-based diagnostic test	Replace standard culture methods to screen & diagnose CPE	Developed and CE-marked	Current culture-based method of detecting CPE	
Oxford	COPD patients with acute exacerbation	Primary care	RightStart	Distinguish between inflammatory and bacterial exacerbation to guide proper medication and reduce unnecessary antibiotic prescription	Under development	Clinical judgement	(10)

Experiences

Our experiences can be grouped into three core areas that summarise the ways in which early modelling of diagnostics differs from late cost-effectiveness analysis:

1. Early stage of test development
2. Uncertain clinical pathway
3. Limited evidence

In each of these areas, early modelling was associated with a number of strengths and challenges.

1. Early stage of product development

Strengths

The early development and evaluation status of a diagnostic technology in the context of early modelling proved to be an advantage for the collaboration of both developers and researchers. Predominantly, this came in the form of a willingness for test developers to be responsive to the results of the analysis. There was strong buy-in from developers, who used the results to inform further development of the technology and ensure the design of devices was fit for purpose within appropriate environments of use. At an early stage, results can be used not only to decide whether to progress with or discard a technology, but also to investigate different clinical applications and patient populations before prioritising the evaluation of a particular application that is likely to bring most value to various healthcare environments. Hence, researchers felt less pressured to have the 'right' outcome and, instead, shared possible outcomes which provided developers with a better context for decision-making in order to ensure the test they develop is fit for purpose.

Researchers also felt at greater liberty to investigate the potential added value of the technology when taking a broader perspective than that of the health system, such as including societal costs and outcomes. This additional information, although not routinely considered in later Health Technology Assessment (HTA), could provide additional insights into potential mechanisms of benefit to inform further R&D. For example the OPTIMAprelim model helped us to understand that capturing the effect of the test on employment might affect a proportion of the patient population (9).

Challenges

Although the models and research teams were funded independently of their industry collaborators, reducing pressure to find the "right" results, working with companies whose device development process was high-risk remained a challenge. Compared with pharmaceutical companies, device developers are often small companies with very few devices. Smaller companies are at risk of being bought out or losing members of staff to other ventures. Hence, there can still be pressure to produce evidence that supports financial stability and company growth. This lack of size and stability, alongside the time constraints imposed by this particular funding call, also increased the need for results to be produced rapidly, which can conflict with the need to invest significant time in model development to produce valid and robust results.

2. Uncertain or flexible clinical pathway

Strengths

Introducing a diagnostic into a pathway where none previously existed, as many of these technologies intend, gives rise to a host of challenges when trying to map out the clinical care pathway and establish the consequences of testing. Measures of clinical utility require evidence of treatment effectiveness in patients who receive both accurate and inaccurate test results. These cannot be ascertained if previously no test was used. Additionally, although a reference test or comparator may exist, it may not have been used in the target pathway, for example, for reasons of cost, or ethics (if the reference is invasive). A relevant comparator may have been used only in a different setting, such as when tests are transposed from secondary care to primary or community care. The second panel on cost-effectiveness in health and medicine recommends that all relevant comparators be included, including those that are not currently used in the care pathway under evaluation (11). As a result, the consequences of introducing a new test are difficult to benchmark to and highly uncertain. Collaboration between developers, research teams and specialist clinical teams when grappling with these issues is essential.

Clinical expert opinion was used to identify potentially appropriate strategies of the diagnostics, in terms of the clinical decision point (when it is most cost-effective to test), patient subgroup (in whom), and re-testing options (how often, such as when the biomarker changes over time or in response to treatment). For example, for the thromboelastography (TEG)-like device, NICE had identified applications in cardiac surgery, trauma and obstetrics, but clinical expert opinion identified further applications in intensive care units, elderly care and general surgery.

Expert opinion can also be used to refine the strategies modelled. For example, in the COPD model the clinical value of early detection of acute exacerbations in primary care was considered to be low by the general practitioners consulted. Instead, the modelling focussed on treatment stratification, where opportunities to both improve health outcomes and reduce antibiotic prescribing were identified.

Use of clinical opinion to refine potential test strategies is crucial because the computational and evidence requirements necessary to evaluate all possible options would be overwhelming in most cases (although this has been tried previously (12)).

Possible barriers to adoption beyond clinical utility were also highlighted, such as work flow issues and ergonomics. Human factors research was used in the case of the TEG-like device to explore how testing is performed in this hospital setting, with the aim of informing device design. In one case (CPE testing), early identification of potential barriers to adoption informed not only future product development, but also the company business plan and marketing strategy. Again, this is outside the realm of later HTA, but was able to provide support that helped ensure the developer of a cost-effective technology remains solvent through product development and manufacture.

In aiming to provide information that guides technology development, in addition to informing stop-go decisions, early modelling allows flexibility. Where the value of a technology has not been fully characterised, there is space to explore a range of outcomes, both interim and final, according to test developers' and decision makers' evidence needs at the time of the analysis. While cost-utility analysis remains important in this area, researchers also used cost-consequence and budget impact analyses where these were relevant.

For example, in the case of CPE testing, the model was designed to estimate the proportions of the patient cohort that would be identified as true or false positive, and true or false negative, by each method of testing. The company involved in developing the diagnostic test found it useful to consider not only expected costs, but also diagnostic outcomes, and to consider the impact that

prevalence of the condition and test accuracy could have on these outcomes. This allowed exploration of the role of screening in differing regional scenarios where prevalence is likely to be higher than the current national average.

Challenges

Most projects received some clinical support to inform the structure of the model as well as the validation of the completed model, but this was usually limited to local experts due to time and resource constraints. This may limit the applicability of the evaluations to the NHS as a whole, particularly in areas where national guidance does not exist, and as a result local care pathways may vary. In some cases, for example where delayed prescribing is used, this could increase structural uncertainty in the model, in addition to parameter uncertainty. This was noted in the COPD project, where antibiotic prescribing guidelines for COPD are set by local CCG, rather than co-ordinated nationally.

Where best practice guidelines do exist they may not reflect actual clinical practice, particularly as NHS Trusts have differing priorities. This may have in fact provided the motivation for developing a new diagnostic, if the existing one has a high patient burden and poor implementation as a result. Furthermore, access to care pathway evidence sources such as Map of Medicine, a resource providing access to clinical care pathways at the point of care, is limited and requires funding (13). In addition, data about the accuracy of the test under evaluation – when available – were often from studies conducted outside the UK, with different prevalence and incidence rates.

Finally, in the case of the interaction between tests and treatments used in sequence (for example if a rule-in test is followed by a rule-out test), tests are frequently only evaluated in isolation. There is a lack of established methods for modelling this interaction (14).

3. Limited evidence base

Strengths

There is no reason why an evaluation of a test at a single decision point in the care pathway cannot be applied to any other test that could be used at this decision point, with minor adjustments for accuracy, cost and adverse events from testing. Given the prevalence of 'fast-followers' in this segment of biotechnology, the ability to use established models to evaluate new tests, or even provide generic benchmarks for cost and accuracy, are clear advantages of decision modelling and a way of reimbursing the additional time spent creating often-complex models. A more fully developed model also opens up opportunities for exploring uncertainty through sensitivity analysis. The more parameters (such as potential outcomes) are effectively included, the more their contribution to cost-effectiveness can be explored. This proved to be particularly informative in the COPD model, where the main recommendation – to generate evidence of clinical utility via a trial – was entirely informed by the results of two-way sensitivity analysis.

Another advantage is the ability to refine and update these more flexible models as evidence becomes available. This could enable greater evidence generation efficiency in later stages of development.

Regarding the method of evaluation, investment in a comprehensive model could facilitate advanced analyses such as value of information (VOI) analysis. VOI is an increasingly popular tool

that can be used to generate information regarding the value of further investment into research, providing information that complements sensitivity analysis (15). Within the early modelling context, VOI can be used to identify key parameters in the model that are driving uncertainty around the expected cost-effectiveness outcomes, and thereby direct the focus and design of future research and development activities (16,17). VOI formed the basis of the recommendations of two projects outlined here (OPTIMA and AKI).

Challenges

The downside of developing flexible, iterative models that follow a test through its development is that well-executed early models are not necessarily faster, easier to implement, or less complex than models carried out at the late stages of technology development. This is unsurprising for two reasons. First, for VOI and effective sensitivity analysis to be conducted, a model needs to describe an intervention's effects on outcomes as accurately as possible. Second, if the intention is for the model to be iteratively developed and reanalysed throughout the pipeline as new evidence becomes available, then for this process to be efficient, the early model must resemble as closely as possible the late model it will eventually become.

A further challenge is the limited data not just for populating the model, but also to fully characterize the technology, for example, shelf life and ease of use, which may affect implementation. Flexibility is both an advantage and a challenge for evaluators using the model results, as there are many more possible scenarios to model.

It is also difficult to validate such models, particularly in relation to current treatment effects that are informed by a large number of assumptions. The plausibility of these assumptions were not always easy to assess. For example, if a patient population has never been differentiated by testing before, because no diagnostic exists, then the difference in average outcomes in the two groups is almost impossible to model, even though it may be critical for determining test cost-effectiveness.

Within the conventional cost-utility framework used in economic evaluation, diagnostic technologies face an additional hurdle compared with treatments. In order to demonstrate an impact on clinical outcomes, a diagnostic must change clinical decision making. However, clinicians and patients frequently overrule test results, particularly in primary care, with the result that test cost-effectiveness is reduced. This aspect of clinical and patient decision making is either ignored or often overlooked in its impact on cost-effectiveness, with the assumption being that technologies are perfectly implemented and results are fully followed.

Methodological Developments Needed

Our experiences have highlighted a number of limitations in current methods and shortfalls in the available evidence which require addressing in order to fully realise the potential of early modelling for diagnostic development and evaluation. These cover clinical pathway mapping, model transparency, and modelling sequential testing.

Pathway mapping

Most significantly, modelling diagnostic tests requires a comprehensive understanding of how a test will alter patient care. Diagnostics have a largely indirect effect on health outcomes, and consequently evidence is needed to understand the full scale of the effects of testing on health

outcomes and costs. While some pathways, such as the bowel cancer screening programme, have clear guidance on treatment following testing (and, therefore, a mechanism of benefit), other clinical areas lack this. For example, the COPD exacerbation management pathway in primary care was found to be highly variable, based on evidence from national and international guidelines, epidemiological evidence, and clinical expert opinion. Notably, which patients received antibiotics, and how they were followed up and, if necessary, given further treatment was unclear and highly variable between evidence sources. In this case, currently available evidence, even based on routinely collected data, did not fully capture current clinical practice, and understanding and parameterizing the likely benefit of testing was thus more difficult. The solution was to include the uncertainty in the model, and when this uncertainty affected cost-effectiveness, use this to support additional primary research into clinical decision making in this care pathway, which is ongoing.

A number of the projects addressed this issue by using extensive care pathway mapping exercises, usually involving expert elicitation (18). All the projects took a slightly different approach to this mapping, depending on both the setting in question and the methodological preferences of the researchers. A valid, consistent approach to the pathway mapping process that provides guidance on combining published evidence with expert opinion would improve this.

Model transparency

Another improvement would be greater transparency around models. Model registries, where all the information necessary to replicate and critically appraise a model is published online, are of increasing interest (19). The presence of model registries and open models could increase the efficiency of projects and provide a path to greater consistency between evaluations. Model building is a resource intensive exercise. It relies on a substantial number of assumptions, and when considering multiple models of interventions in the same care pathway, relies on an assumption of commonality: that a treatment that is cost-effective would fundamentally be so in all models, and vice versa. Registries and open, or shareable, models would improve this commonality.

Registries and greater sharing would also enable use of more specialised model structures, such as dynamic transmission models for infection – assuming the model documentation is also transparent and comprehensible. Two of the models developed were for tests that were expected, and intended, to have some impact on antimicrobial resistance (AMR). Modelling the impact on AMR is likely to be an increasing requirement for diagnostics modelling, since the publication of the O'Neill report — a UK government-commissioned report that explored the likely global health and economic consequences of AMR — concluded that all antibiotic prescribing should be preceded by testing (20). The need to include AMR in economic evaluation is recognised, but methods for doing so are not yet well-developed (21).

Sequential tests

Finally, none of the care pathways modelled here relied on a single test to exclusively guide treatment, but diagnostic evidence overwhelmingly evaluates tests in isolation, compared to a reference standard. The assumption that there is independence between multiple tests used may overestimate the diagnostic accuracy of the pathway (22).

Ultimately, these issues affect decision modelling at all stages of product development, and are not limited just to diagnostics. However, their presence is felt acutely in early diagnostic modelling, where the issues compound in the absence of the cushion of better evidence.

Outstanding questions around how and when early models should be used for diagnostics

Although some of the challenges outlined above can be addressed through better evidence and methods development, others raise normative questions around the purpose of modelling.

Although early economic modelling can be useful to inform the key decisions in the development of a new technology, caution is required when making decisions on the basis of inevitably weak or limited evidence. This raises the question of what the minimum evidence requirements are to build an early economic model and what decisions can be made on the basis of this exercise. Most of the companies we worked with had at least some data on test accuracy, but none had information on the downstream consequences of the resulting treatment decisions and how testing may affect these. Modelling can fill this gap if other evidence sources on treatment effectiveness are available, but this evidence was not always available, and the regulatory and funding landscape does not appear to be in place to incentivise its generation, given the overwhelming focus in the IVD industry on CE marking and market authorisation based solely on diagnostic accuracy (2).

There is also a balance to be struck between complex models that can be adapted over time, and efficient, simple models that can answer specific questions rapidly. In some cases, the tests under evaluation could be of potential value in multiple care pathways, or at multiple points in the same care pathway. For example, the imaging agent under evaluation had potential value within a bowel cancer screening programme, and also in patients under surveillance for IBD and Lynch disease. Models for multiple care pathways were built, but this was challenging given the time-frame and financial constraints. It may have been more pragmatic to do a very simplistic decision tree analysis for each pathway initially to narrow down the focus of the early economic analysis, but the validity of such an approach is unclear.

Finally, as access to linked big data slowly becomes more of a reality, electronic healthcare records could be used to give a more realistic 'real world' clinical pathway to inform the structure of early economic models. However, the risk is that this will result in a very complicated analysis, as patients rarely conform to a single 'recommended' clinical pathway. This has recently been demonstrated by a recent mining exercise to determine real-life clinical pathways during chemotherapy where 474 different pathway variants were identified among 535 patients with breast cancer (23). That said, if current models of care pathways are not sufficiently representing the complexity of real-world clinical practice, it could be argued that neither are the results of our analyses.

Advice for researchers considering early modelling for tests

Based on these experiences, existing good practice guidelines, and discussions we have had in developing this paper, we have outlined a number of recommendations for health economists and their collaborators when developing early models of diagnostics (table 2).

1. *Establish the key questions to be answered*

This is an established requirement of economic models generally (24,25). However, in an early model the range of decision problems to be addressed may be broader, including risks to continuing development, headroom in terms of diagnostic accuracy or price, and budget impact.

If a simplified modelling approach is necessary due to a significant lack of relevant evidence, it can still be informative and useful from the developer's perspective, this was seen in the case of CPE

testing. Relevant early information gathering, even if only informally, for example on the burden of testing on patients or clinicians, can identify risks and barriers to adoption of the technology at an early stage.

To facilitate this, the key purpose of the model should be addressed before embarking on its development. This will enable identification of the relevant outcomes, time horizon and perspective to consider, as well as the most appropriate analysis to use e.g. headroom analysis, cost-consequence, effectiveness or utility analysis.

2. *Develop a model that reflects of the care pathway*

A fully developed model can be developed and updated alongside the research being carried out over time. Often, lack of data availability and time pressures at early stages of product development result in a simplified approach to modelling. However, early investment in a model that can capture all the components of a test / treat pathway as accurately as possible will increase efficiency later in the process, as the model will have the scope to absorb new evidence without the need for complete redevelopment.

An early model can easily be misinterpreted as a 'quick and dirty' analysis, but in fact may be more valuable when it is a full analysis which can be iteratively updated as further evidence becomes available. As outlined in the previous section, while diagnostic accuracy is often assumed to be the key parameter in diagnostic models (Ferrante di Ruffano, BMJ 2012), there are other ways in which diagnostic technologies can lead to downstream changes in patient health. This finding was reflected in our experience where linking test results to clinical outcomes was found to be a critical determinant of cost-effectiveness in almost all cases.

Of course, funding is not unlimited and research resources must be prioritised, but the long term efficiency that could result from early and iterative model development could make this an appealing goal for funders. Where a full model is not possible, establishing key questions as outlined above will guide the most effective alternative analytical approach.

3. *Early and frequent stakeholder engagement*

Early engagement with relevant stakeholders e.g. clinicians, laboratory managers, policy makers, and patients can ensure that the care pathway modelled adequately reflects current practice, and the new pathway proposed is plausible, acceptable and generalizable. Stakeholders can also provide insight into the limitations of a plausible model. Consideration of expert opinion can also help identify the relevant outcomes (e.g. health and cost outcomes, and NHS resource use) likely to be affected by the introduction of the new technology, and therefore help to identify the optimal role(s) and setting(s) of the new test, which may directly inform product requirements and, therefore, future development.

If possible, continued engagement with clinical stakeholders throughout model development and analysis can ensure 'sense checking' of the inputs and outputs of the model. An added advantage of engagement with relevant stakeholders is that barriers and facilitators to implementation can be identified and enable the developer to take steps to overcome them.

4. *Use adaptive review methods*

Systematic review methods are commonly used to inform model structure and quantify model parameters in rigorous health economic modelling studies. However, these approaches may be too resource intensive for models at an early stage, especially if initial searches reveal a limited

evidence-base, as was the case with the COPD project. Here, rapid review methods proved more efficient.

Accordingly, the majority of projects used targeted literature searches, where searching and screening were conducted iteratively, rather than following a fixed protocol, to allow for flexibility. This was combined with stakeholder consultation to rapidly inform model structure and parameterisation. It is important to allow enough time in the model building process to supplement searching with expert elicitation and stakeholder consultation, to fill in the gaps.

5. *Include meaningful sensitivity analysis*

Uncertainty in both model structure and input parameters is inevitable in decision modelling. Early modelling of diagnostics amplifies this uncertainty, first through lack of evidence, and second through the additional layer of uncertainty required in modelling diagnosis *and* treatment. Conducting deterministic two- and three-way sensitivity analysis over plausible values of input parameters can identify potentially important thresholds, such as the maximum price of a test, the minimum necessary diagnostic accuracy, and the interaction between the two. The results of such an analysis are likely to be more meaningful than the headline ICER, by providing evidence for decisions, such as price, that are still within a developer's control. The outcomes of deterministic sensitivity analysis in the COPD project also emphasised the need for treatment effectiveness evidence, the strongest recommendation provided by that research.

Probabilistic sensitivity analysis has value in assessing the impact of the overall uncertainty of the model parameters on model outputs, although it is likely to simply (and inevitably) tell us that cost-effectiveness is highly uncertain. It is more useful when used as the basis for VOI to inform future research. Probabilistic sensitivity analysis was the basis of the VOI in both the AKI and OPTIMA models and formed the basis of research recommendations. An early sensitivity analysis plan may prove useful in ensuring the most relevant information is interrogated.

Table 2. Key advice for researchers conducting early models of diagnostics

1	Establish the key questions that need to be answered to inform future development of the test
2	Aim to develop an iterative model that can be adapted over time, improving future research efficiency
3	Early engagement, particularly with independent clinical experts and decision makers
4	Use adaptive review methods and allow time to supplement the literature with expert opinion
5	Include meaningful sensitivity analysis that will answer the key questions, including value of information analysis where possible.

CONCLUSION

Diagnostic tests are a pivotal component of the much-mooted paradigm shift towards personalised healthcare. They are also complex to research, develop, and evaluate, with high barriers to adoption. As such, we believe that more intelligent approaches to research and development for tests are necessary for the realisation of the benefits of test-based decision making. The NIHR DECs have demonstrated the usefulness of early consideration of the impact of a test on clinical pathways, and its likely implications for clinical and economic outcomes. We believe this approach should be part of

the core strategy in the movement towards more efficient research design and timely delivery of high value and evidence-based diagnostic products for patient care.

Acknowledgements

We would like to thank the attendees of the 2018 Winter HESG meeting for their advice on an early draft of this paper. We also acknowledge the support of our industry collaborators in the early modelling projects that led to this work.

Conflicts of interest

The authors declare that there is no conflict of interest.

References

1. Bossuyt PMM, Reitsma JB, Linnet K, Moons KGM. Beyond diagnostic accuracy: the clinical utility of diagnostic tests. Clin Chem [Internet]. 2012 Dec [cited 2019 Mar 27];58(12):1636–43. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22730450

2. Verbakel JY, Turner PJ, Thompson MJ, Plüddemann A, Price CP, Shinkins B, et al. Common evidence gaps in point-of-care diagnostic test evaluation: a review of horizon scan reports. BMJ Open [Internet]. British Medical Journal Publishing Group; 2017 Sep 1 [cited 2017 Nov 21];7(9):e015760. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28864692

3. Annemans L, Genesté B, Jolain B. Early Modelling for Assessing Health and Economic Outcomes of Drug Therapy. Value Heal [Internet]. Wiley/Blackwell (10.1111); 2008 Nov 1 [cited 2018 Aug 14];3(6):427–34. Available from: http://linkinghub.elsevier.com/retrieve/pii/S1098301511700122

4. Buisman LR, Rutten-van Mölken MPMH, Postmus D, Luime JJ, Uyl-de Groot CA, Redekop WK, et al. The early bird catches the worm: early cost-effectiveness analysis of new medical tests. Int J Technol Assess Health Care [Internet]. Cambridge University Press; 2016 Mar 22 [cited 2016 Oct 7];32(1–2):46–53. Available from: http://www.journals.cambridge.org/abstract_S0266462316000064

5. National Institute for Health Research. Diagnostic Evidence Co-operatives. 2017.

6. Randox [Internet]. 2016 [cited 2018 Jul 24]. Available from: https://www.randox.com/tag/bladder-cancer/

7. Sutton A, Lamont J, Evans M, 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;[In press].

8. Hall PS, Mitchell ED, Smith AF, Cairns DA, Messenger M, Hutchinson M, et al. The future for diagnostic tests of acute kidney injury in critical care: evidence synthesis, care pathway analysis and research prioritisation. Health Technol Assess (Rockv). 2018 May;22(32):1–274.

9. 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 Heal J Int Soc Pharmacoeconomics Outcomes Res. United States; 2017 Dec;20(10):1311–8.

10. Abel L, Dakin HA, Roberts N, Ashdown HF, Butler CC, Hayward G, et al. Is stratification testing for treatment of chronic obstructive pulmonary disease exacerbations cost-effective in primary care? an early cost-utility analysis. Int J Technol Assess Health Care [Internet]. Cambridge University Press; 2019 Mar 4 [cited 2019 Mar 14];1–10. Available from: https://www.cambridge.org/core/product/identifier/S0266462318003707/type/journal_article

11. Owens DK, Siegel JE, Sculpher MJ, Salomon JA. Designing a cost-effectiveness analysis. 2nd ed. Neumann PJ, Ganiats TG, Russell LB, Sanders GD, Siegel JE, editors. Cost-Effectiveness in Health and Medicine. Oxford University Press; 2016.

12. Mant J, Doust J, Roalfe A, Barton P, Cowie MR, Glasziou P, et al. Systematic review and individual patient data meta-analysis of diagnosis of heart failure, with modelling of implications of different diagnostic strategies in primary care. Health Technol Assess [Internet]. 2009;13(32):1–207, iii. Available from: http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=19586584&retmode=ref&cmd=prlinks%5Cnpapers3://publication/doi/10.3310/hta13320

13. Stein M. The Map of Medicine® - an Innovative Knowledge Management Tool. AMIA Annu Symp Proceedings. 2006;1196.

14. Annemans L, Redekop K, Payne K. Current Methodological Issues in the Economic Assessment of Personalized Medicine. Value Heal [Internet]. Elsevier; 2013 Sep 1 [cited 2018 Aug 16];16(6):S20–6. Available from: https://www.sciencedirect.com/science/article/pii/S1098301513018640?via%3Dihub

15. Ginnelly L, Claxton K, Sculpher MJ, Golder S. Using Value of Information Analysis to Inform Publicly Funded Research Priorities. Appl Health Econ Health Policy [Internet]. Springer International Publishing; 2005 [cited 2017 Dec 11];4(1):37–46. Available from: http://link.springer.com/10.2165/00148365-200504010-00006

16. Hall PS, Mitchell ED, Smith AF, Cairns DA, Messenger M, Hutchinson M, et al. The future for diagnostic tests of acute kidney injury in critical care: evidence synthesis, care pathway analysis and research prioritisation. Health Technol Assess (Rockv) [Internet]. 2018 May [cited 2018 Aug 29];22(32):1–274. Available from: http://www.ncbi.nlm.nih.gov/pubmed/29862965

17. Claxton KP, Sculpher MJ. Using Value of Information Analysis to Prioritise Health Research. Pharmacoeconomics [Internet]. Springer International Publishing; 2006 [cited 2018 Aug 29];24(11):1055–68. Available from: http://link.springer.com/10.2165/00019053-200624110-00003

18. Bojke L, Claxton K, Bravo-Vergel Y, Sculpher M, Palmer S, Abrams K. Eliciting distributions to populate decision analytic models. Value Heal J Int Soc Pharmacoeconomics Outcomes Res. United States; 2010 Aug;13(5):557–64.

19. Sampson C. Call for a model registry [Internet]. The Academic Health Economist's blog. 2012 [cited 2017 Nov 6]. Available from: https://aheblog.com/2012/10/19/call-for-a-model-registry/

20. O'Neill J. The Review on Antimicrobial Resistance. Final report. [Internet]. 2016. Available from: http://amr-review.org/Publications

21. Coast J, Smith R, Karcher A-M, Wilton P, Millar M. Superbugs II: how should economic evaluation be conducted for interventions which aim to contain antimicrobial resistance? Health Econ [Internet]. John Wiley & Sons, Ltd.; 2002 Oct 1 [cited 2017 Sep 25];11(7):637–47. Available from: http://doi.wiley.com/10.1002/hec.693

22. Sutton AJ, Cooper NJ, Goodacre S, Stevenson M. Integration of meta-analysis and economic decision modeling for evaluating diagnostic tests. Med Decis Making [Internet]. 2008 Jan 1 [cited 2015 Oct 26];28(5):650–67. Available from: http://mdm.sagepub.com/content/28/5/650

23. Baker K, Dunwoodie E, Jones RG, Newsham A, Johnson O, Price CP, et al. Process mining routinely collected electronic health records to define real-life clinical pathways during chemotherapy. Int J Med Inform [Internet]. 2017 Jul [cited 2017 Nov 28];103:32–41. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28550999

24. Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D, et al. Consolidated Health Economic Evaluation Reporting Standards (CHEERS)--explanation and elaboration: a report of the ISPOR Health Economic Evaluation Publication Guidelines Good Reporting Practices Task Force. Value Heal J Int Soc Pharmacoeconomics Outcomes Res. United States; 2013;16(2):231–50.

25. Philips Z, Ginnelly L, Sculpher M, Claxton K, Golder S, Riemsma R, et al. Review of guidelines for good practice in decision-analytic modelling in health technology assessment. Health Technol Assess (Rockv). 2004;8(36).