Features, process and methods of early health technology assessment to inform developers of health technologies: a proposed framework and application to diagnostic technologies

Janet Bouttell

Thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

Health Economics and Health Technology Assessment
Institute of Health and Wellbeing
University of Glasgow

15 February 2021

Summary

Key points

- This thesis aimed to develop a framework to help people undertaking health technology assessment of health technologies in development (termed 'analysts') approach their work.

- It used a literature review and learning from five case studies to develop a framework which described the features of this type of health technology assessment, a generic process for undertaking the work and described which methods might be appropriate. Although the case studies are all diagnostic technologies, the framework would be appropriate for those working with small and medium-sized developers of any type of health technology.

- Analysts need to be aware that HTA for developers of health technologies and their investors (termed 'developers') is different to HTA undertaken later in the technology's lifecycle. Compared to HTA undertaken at market access stage, HTA undertaken to inform developers (termed 'development-focused HTA' or 'DF-HTA') informs a broader range of decisions and has to deal with a greater level of uncertainty, as the precise use of the technology may not yet be known and there is likely to be little evidence specific to the technology.

- Although the process of DF-HTA will be familiar to analysts in that it involves the assessment of clinical and economic value, DF-HTA is more iterative in nature and does not have strict methodological guidelines to follow, as is often the case for HTA at market access stage. Analysts can take a more positive stance and put themselves in the shoes of the developer to consider what analysis would be useful to the developer at this stage of development.

- The methods would be familiar to those carrying out health technology assessment but need some modification when the work is to inform developers. This is particularly relevant when resources for analysis are

constrained, which is often the case when assessing technologies in development, many of which are likely to fail. In practice, this means that methods may be simpler. There is also more emphasis on methods of stakeholder consultation as a result of the lack of clinical evidence.

- The framework would benefit from validation through prospective use in further case studies by external groups of analysts. This is particularly important because much HTA work undertaken for commercial developers is not published due to commercial sensitivity.

Background and aims

The development of health technologies is expensive and high risk, as many technologies in development fail to achieve commercial success. There are many choices to be made during the development process and the decisions made at this time are likely to influence the future success of the technology. It has been suggested that health technology assessment (HTA) undertaken during the development process may help to reduce the risk of failure or to accelerate failure and reduce research waste. The form of HTA used to inform developers about some of the decisions they need to make during the development process has been termed 'Early HTA'. However, the current 'Early HTA' literature includes work which is not intended to directly inform developers (such as HTA accompanying managed access agreements) and much of the HTA that is conducted to inform developers is not published. As a result, there is a lack of clear guidance for the analyst on how to approach HTA to inform developers of medical technologies.

This thesis proposes a framework to aid analysts (people undertaking HTA assessments) who are working with developers of medical technologies. The term developers is used throughout to describe the individuals or companies who are developing a medical technology and their investors/funders. Large pharmaceutical and medical device companies tend to have in-house HTA resource (often termed Health Economics and Outcome Research (HEOR) or market access teams). For this reason, the framework is aimed primarily at academic and consultant analysts working with small or medium-sized enterprises (SMEs).

The framework comprises three sections: features, process and methods of HTA to inform developers (termed development-focused or DF-HTA). Features means the characteristics of DF-HTA. Process describes the broad activities of DF-HTA. Methods describes the analytical approach to those activities or how the analyst would undertake the activity.

The framework was developed alongside five case studies in the assessment of diagnostic technologies which are included to illustrate the framework of features, process and methods.

Methods

A pearl-growing literature review was undertaken in October 2017 and refreshed in February 2019. Pearl-growing involves searching the citations and references of articles of interest until no further papers are being retrieved. The aim was to retrieve literature which informed an understanding of early health technology assessment and included both methods papers and applied studies.

An iterative process was undertaken to develop a framework of DF-HTA including features, process and methods. An initial list of features of DF-HTA was developed then refined and expanded using an iterative process informed by the content of articles identified as being particularly informative and the experience of undertaking the case studies. Frameworks for the process of DF-HTA identified in the literature review were used as the basis for the development of a generic process of DF-HTA. Methods useful for DF-HTA were extracted from review articles. Methods of research and development or other commercial activities were excluded. Terminology was streamlined where similar terms were describing the same method.

Five case studies were used to illustrate the framework of DF-HTA. For each of the case studies the features were compared to the list of features and areas of non-conformity identified and discussed. The process of DF-HTA included clinical value assessment in all cases and economic value assessment in four of the five case studies. Methods were selected from the methods identified as useful for DF-HTA.

Results

A total of 152 early HTA papers were identified of which 88 were judged to be aimed at informing developers. These comprised 56 methods, 61 applied and 35 methods and applied papers in early HTA of which 43 methods, 25 applied and 20 methods and applied were aimed to inform developers.

A proposed framework of DF-HTA was developed including the features, a generic process and a range of methods suitable for DF-HTA. Ten features characterising DF-HTA were identified including six which had not been previously identified: audience; timing; decisions informed; available evidence; underlying user objective; decision space; business model; resources for analysis; stance of analysis and burden of proof. The proposed generic process of DF-HTA comprises two core aspects of DF-HTA identified in all the frameworks found in the literature – clinical value assessment and economic value assessment. Clinical value assessment considers what impact the technology might have on clinical practice and ultimately upon health (and wider social) outcomes. Economic value assessment builds on the clinical value assessment to consider the economic impact of changes in healthcare resource use and other economic value drivers such as productivity effects. Eight broad groups of methods useful in DF-HTA were identified from review articles: care pathway analysis; qualitative methods of stakeholder interaction; literature review; multi-criteria decision analysis; discrete choice experiments; expert opinion and elicitation; cost effectiveness analysis and value of information analysis.

The framework was illustrated using the case studies. The features of DF-HTA were evident in three case studies of technologies in development but not in two case studies concerning technologies already in clinical use. The process and methods used in the case studies aligned with the generic process of DF-HTA. All the case studies began with a clinical value assessment using decision models to map the existing and potential clinical pathways. For the three DF-HTA case studies parameter estimates came from the literature and expert opinion initially supplemented by evidence from small clinical studies specific to the technology where this was available. The case studies show that an initial clinical model is able to distinguish technologies with potential value from those with little potential value. Cost-effectiveness or cost consequence models were

able to identify potential economic value, indicate design factors which required clarification before further development and identify parameters likely to have a significant impact on potential economic value. The two case studies which involved technologies which were already in clinical use were able to show, using similar methods to DF-HTA and in situations where evidence was limited, cost savings and consequences of expanding a testing programme and introducing a triage test. These case studies demonstrate the need for alternatives to cost-effectiveness using methods appropriate at market access/adoption stage when resources are limited.

Conclusion

The main contribution of the thesis was to propose a framework of features, process and methods of DF-HTA. A secondary contribution was to provide five empirical examples, three in DF-HTA and two in early economic evaluation of diagnostic technologies. The case studies use simple models that can be readily used to provide an indication of the potential value of a technology whilst it is under development or during some form of expedited review.

The main limitation concerns the likely incompleteness and unrepresentativeness of HTA studies to inform developers in the academic literature. This is because of commercial confidentiality and a lack of an incentive to publish.

This thesis suggests that developers should be encouraged to consult an HTA practitioner at an early stage in the development. Where development is technology-driven (i.e. the technology is developed without a precise clinical need identified), DF-HTA can help developers to 'position' their technology, articulate value propositions and engage with stakeholders. This process can sometimes change the design or intended direction of the technology in development and can inform evidence generation strategy. For analysts, the framework should provide guidance on the nature, process and methods of HTA to meet the needs of the developer. For policy-makers the thesis suggests that it is possible and desirable to encourage the articulation of a value proposition (particularly for translational research) within funding applications. It is also essential to support translational research bodies which facilitate links between

commercial entities, academic and clinical researchers, clinicians, regulators and reimbursement agencies. Policy-makers may also wish to fund the development of full disease models which could be used to rapidly and robustly evaluate any proposed technology (in technology-driven development) as well as determining areas of greatest need to inform specific calls for innovation (needs-driven development).

Further research exploring the features of in-house and unpublished DF-HTA and the usefulness of this framework of DF-HTA to developers and analysts would be valuable. The decision-making process for the adoption of many technologies is not clear and transparent and research into the evidence relied upon by decision-makers in different settings would be useful. The extent to which full disease models have been or could be used to assess the value proposition for innovative technologies would also be a useful area of research.

Contents

List of Tables

List of Figures

Acknowledgements

I would like to thank Professors Neil Hawkins and Andy Briggs for their support and advice during the preparation of this thesis. The current and former occupants of Room 202 have been supportive throughout and other friends at HEHTA have contributed as readers, collaborators and providers of moral support when required.

Many clinicians and academics have given their time to the various case studies presented here. I would like to thank all collaborators on papers as well as Iain McNeish, Iain McInnes, Duncan Porter, Carl Goodyear and Tricia Roxborough. The Glasgow Molecular Pathology Node provided the context and funding for this thesis and much food for thought in the many interactions at Node events. It was a pleasure working with Ruth McLaughlin, Karin Oien and the other node staff.

Finally, many thanks to Phil, Freddy, Ellie, Joe, Tom, Mum, Dad, Laura, Jane and all the other friends and family who have patiently listened after asking 'How is the PhD going?'

Publications and presentations

The following publications and presentations are a result of the research conducted for this PhD.

Published

Bouttell, J., Teoh, J., Chiu, P.K., Chan, K.S., Ng, C.F., Heggie, R. and Hawkins, N., 2019. **Economic evaluation of the introduction of the Prostate Health Index as a rule-out test to avoid unnecessary biopsies in men with prostate specific antigen levels of 4-10 in Hong Kong**. *PloS one*, *14*(4), p.e0215279.

Bouttell, J., Tan, Y.Y., Creed, D., McGaffin, G., Hawkins, N., McLaughlin, R., Smith, G., Westwood, P., Williams, N. and Graham, J., 2019. **Context-Specific Economic Evaluation for Molecular Pathology Tests: An Application in Colorectal Cancer in the West of Scotland.** *International Journal of Technology Assessment in Health Care*, *35*(4), pp.327-333.

Bouttell, J., Briggs, A. and Hawkins, N., 2020. **A different animal? Identifying the features of health technology assessment for developers of medical technologies.** *International Journal of Technology Assessment in Health Care, pp.1-7.*

Bouttell, J. , Grieve, E. and Hawkins, N., 2020. **The role of development-focused health technology assessment in optimizing science, technology, and innovation to achieve sustainable development goal 3**. In: Adenle, A. A., Chertow, M. R., Moors, E. H.M. and Pannell, D. J. (eds.) Science, Technology, and Innovation for Sustainable Development Goals: Insights from Agriculture, Health, Environment, and Energy. Oxford University Press: New York, NY. ISBN 9780190949501

Bouttell, J. , Hawkins, N. and Grieve, E., 2021. **Maximizing the value of engineering and technology research in healthcare: development – focused health technology assessment.** In: Imran, M. A., Ghannam, R. and Abbasi, Q. H. (eds.) Engineering and Technology for Healthcare. Wiley-IEEE, pp. 1-27. ISBN 9781119644248

In preparation

Bouttell, J., Hawkins, N., Briggs, A. and Ponomarev, D. **The Test and Treat Superiority Plot: a rapid assessment tool to evaluate 'precision medicine' tests predicting individual response to treatment**.

Bouttell, J. and Hawkins, N. **Evaluation of triage tests when existing test capacity is constrained: application to rapid diagnostic testing in CoVID-19.**

Bouttell, J., Briggs, A. and Hawkins, N. **Horses for courses? Methods of development-focused health technology assessment.**

Presentations

Janet Bouttell, Neil Hawkins, Andrew Briggs and Dmitry Ponomarev. **Test and Treat Superiority Plot: developing a simple tool to estimate required test performance for developers of tests for treatment response**. MEMTAB virtual conference, 10th December 2020

Janet Bouttell, Neil Hawkins, Andrew Briggs and Dmitry Ponomarev. **Test and Treat Superiority Plot: developing a simple tool to estimate required test performance for developers of tests for treatment response**. MEED symposium, University of Leeds, 26th September 2018

Janet Bouttell. **Early Health Technology Assessment. Increasing/optimising the value of Node projects**. Glasgow Molecular Pathology Node Annual Symposium. Queen Elizabeth University Hospital. 7th September 2018.

Janet Bouttell, Neil Hawkins, Andrew Briggs. **Identifying cost effective methods of health technology assessment for developers – the need for fast and frugal evaluation**. Poster presentation at ISPOR conference, November 2017

Abbreviations

5-FU	5-fluorouracil
5-HTTLPR	Serotonin-transporter-linked polymorphic region
ACR	American College of Rheumatology
A&E	Accident and Emergency Department
ADA	Adalimubab
AHP	Analytic Hierarchy Process
AIS	Adolescent Idiopathic Scoliosis
BAP	Bio-artificial pancreas
BCR-ABL	breakpoint cluster region - Abelson murine leukaemia viral oncogene
bDMARD	Biologic Disease Modifying Anti-Rheumatic Drug
BRCA1	breast cancer 1, early onset gene
BRAF	B-Raf Proto-Oncogene
CADTH	Canadian Agency for Drugs and Technologies in Health
CBA	Cost-benefit analysis
CCA	Cost-consequence analysis
CCG	Clinical Commissioning Group
cDMARD	Conventional Disease Modifying Anti-Rheumatic Drug
CE	Conformité Européene
CEA	Cost-Effectiveness Analysis
CI	Confidence Interval
CLIA	Clinical Laboratory Improvement Amendments
CMA	Cost minimisation analysis
COPD	Chronic obstructive pulmonary disease
CRC	Colorectal cancer
CrI	Credible Interval
CT	Chemotherapy
CTA	Constructive Technology Assessment
CTC trap	Circulating Cancer Cell trap
CTLA4	Cytotoxic T-lymphocyte-associated protein 4
CTMM	Center for Translational Molecular Medicine
CTS	Clinical Trial Simulation
CUA	Cost-utility analysis
CVD	Cardiovascular Disease
DAP	Diagnostic Appraisal Programme
DAS28	Disease Activity Score 28 joints
DAS28-ESR	Disease Activity Score 28 joints & Erythrocyte Sedimentation Rate
DCE	Discrete Choice Experiment
DEC	Diagnostic Evidence Cooperative
DF-HTA	Development-focused HTA
DNA	Deoxyribonucleic acid
DRE	Digital Rectal Examination
EBM	Evidence Based Medicine
EGFR	Epidermal growth factor receptor
EGFRi	Epidermal growth factor receptor inhibitors
ENBS	Expected Net Benefit of Sampling
EQ-5D	Euroqol 5 dimensions
ErespA	Expected response from treatment A
ErespB	Expected response from treatment B
ErespTest	Expected response from the testing strategy
ESR	Erythrocyte Sedimentation Rate
ETN	Etanercept
EU	European Union
EUA	Emergency Use Authorisation
EULAR	European League Against Rheumatism
EVPI	Expected value of perfect information
EVPPI	Expected Value of Perfect Parameter Information
EVSI	Expected Value of Sample Information
FDA	Food and Drug Administration
GMP Node	Glasgow Molecular Pathology Node
GP	General Practitioner

HAQ	Health Assessment Questionnaire
HIV	Human Immunodeficiency Virus
HK$	Hong Kong dollar
HLA-B*5701	A genetic variation that is linked to hypersensitivity to the antiretroviral (ARV) drug abacavir
HT	Hormone Therapy
HTA	Health Technology Assessment
ICER	Incremental Cost Effectiveness Ratio
ICER	Institute for Clinical and Economic Review
ISD	Information Services Division (of the NHS in Scotland)
ISPOR	International Society for Pharmacoeconomics and Outcomes Research
IV	intravenous
IVDR	In vitro devices regulation
IVI	Infusional intravenous
JAK	Janus kinase
KRAS	Kirsten Rat Sarcoma
LDT	Laboratory Developed Test
MATCH	Multi-disciplinary Assessment of TeCHnologies
MCDA	Multi-criteria Decision Analysis
MDM	Multi-disciplinary team meeting (also sometimes referred to as MDT or MDTM)
MDR	Medical Devices Regulation
MeSH	Medical Subject Headings
MHRA	Medicines and Healthcare products Regulatory Agency
MIC	MedTech and In Vitro Diagnostics Co-operative
MM	Malignant melanoma
ml	Millilitre
MPEP	Molecular Pathology Evaluation Panel
MRI	Magnetic Resonance Imaging
MSc	Masters of Science
MTA	Medical technology assessment
MTX	Methotrexate
n	Number
N/A	Not applicable
NFS	Non-fusion surgery
ng	Nanogram
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
NIHR	National Institute for Health Research
NMB	Net monetary benefit
NRAS	Neuroblastoma RAS viral oncogene homolog
OC	Ovarian Cancer
OR	Odds Ratio
OWSA	One-way sensitivity analysis
pA	Probability of response to treatment A
PAM	Photo-acoustic mammography
pB	Response to treatment B
PBAC	Pharmaceutical Benefits Advisory Committee
PCa	Prostate Cancer
PDL1	Programmed death ligand 1
PEST	Political Economic Social Technological
PhD	Doctor of philosophy
PHI	Prostate Health Index
PICC	Peripherally inserted central venous catheter
PICO	Population Intervention Comparator Outcome
PMA	Pre-market approval
POC	Point of care
PPV	Positive predictive value
PSA	Probabilistic Sensitivity Analysis
PSA	Prostate Specific Antigen
PSS	Personal Social Services
QALY	Quality Adjusted Life Year
RA	Rheumatoid Arthritis
R&D	Research and development
RAS	Rat Sarcoma
RCT	Randomised Controlled Trial

RECIST	Response evaluation criteria in solid tumours
resp	Response
Resp|Test+ve	Response conditional upon test result being positive
Resp|Test-ve	Response conditional upon test result being negative
RNA	Ribonucleic Acid
ROC	Receiver Operator Curve
ROI	Return on Investment
RTX	Rituximab
SCBR	Summary Clinical Benefit Rate
sens	sensitivity
SHELF	Sheffield Elicitation Framework
SIGN	Scottish Intercollegiate Guideline Network
SMART	Simple Multi-Attribute Rating Technique
SMC	Scottish Medicine Consortium
SME	Small and medium sized enterprises
spec	Specificity
SWOT	Strengths Weaknesses Opportunities Threats
TAP	Technology Appraisal Programme
TCZ	Tocilizumab
TNFi	Tumour necrosis factor inhibitor
TPMT	Thiopurine methyltransferase or thiopurine S-methyltransferase
TRUS	Trans-rectal ultrasound-guided
TTS plot	Test and Treat Superiority Plot
UK	United Kingdom
US	United States of America
US$	US dollar
VOI	Value of Information
WGS	Whole genome sequencing
WoSCAN	West of Scotland Cancer Network

1 Introduction

A group of privately-funded, UK-based, developers of a diagnostic technology in malignant melanoma are currently making decisions about whether to continue development of their technology and, if so, how to proceed. One clinical study is complete and several more are in the planning stage. They are considering commissioning some health technology assessment (HTA) to demonstrate to healthcare decision-makers in the UK that the technology has the potential to be cost-effective. Before cost-effectiveness can be addressed, the developers need to consider where the technology could be placed in the clinical pathway and what difference the technology could make to costs, health outcomes and other dimensions of health-care delivery. With little clinical evidence to go on, how could an HTA analyst begin to help developers to answer these questions? Where can the analyst look for guidance as to how to undertake this assessment?

1.1 Motivation

This thesis was undertaken in the context of the Glasgow Molecular Pathology Node (GMP Node); a Medical Research Council/Engineering and Physical Sciences Research Council funded body charged (inter alia) with accelerating the translation of molecular pathology technologies. The GMP Node brought together industry partners, Glasgow and Greater Clyde National Health Service (NHS) and the University of Glasgow. The developers engaging with the GMP node were anticipated to be academic groups from within the university or small and medium sized enterprises (SME) industry partners developing molecular diagnostics. The main focus of the thesis is to examine the features, process and methods of HTA which would be useful for all developers of health technologies. However, there is a particular focus on academic and SME developers of diagnostic technologies given the GMP Node context. The initial concept of the thesis was that HTA methodology would be used to prioritise development projects for GMP Node support. However, the number of appropriate projects was, in the event, too limited to allow the thesis to focus on prioritisation but three GMP Node projects were assessed and are included as case studies. A further two case studies of diagnostic technologies were identified from sources other than GMP node.

Developers of health technologies face many choices including: whether and how to develop the technology; where to market it; and, how to price it. Decisions made during the development process are likely to influence the future success of the technology. The development of health technologies is challenging. There are the challenges which face the developer of any technology such as designing a technology which is affordable, which works as planned and for which there is a demand. Over and above these generic challenges, developers of health technologies need to navigate the complex international regulatory and reimbursement environment. Developing medical technologies of all kinds is high risk, as investment costs are high and the chances of any individual technology in development reaching commercial success are low (Lehoux et al, 2008; van Norman, 2016a, 2016b).

HTA is widely used to inform decision making about the usage, reimbursement and adoption of health technologies at the time of initial market access or later in the technology lifecycle (INAHTA, 2020). HTA has also been suggested as a means to inform developers about some of the decisions they need to make during the development process (Pietzsch and Pate-Cornell, 2008). This form of HTA has been termed 'Early HTA' (Ijzerman and Steuten, 2011). An analyst, unaccustomed to undertaking HTA to inform developers of health technologies, may look to the 'Early HTA' academic literature (for example, as defined by Ijzerman et al, 2017) for guidance on their approach and the methods to be adopted.

The 'Early HTA' literature (Ijzerman et al, 2017) is diverse in audience, timing, methods and terminology. As a result, it does not serve as a clear guide to the analyst on how to approach HTA to inform developers of medical technologies. Some earlier studies were clearly focused on commercial developers of pharmaceuticals (for example: Mauskopf et al, 1996; Annemans et al, 2000) and SME developers of medical devices (for example: Cosh et al, 2007; Craven et al, 2009; Chapman, 2013). However, in the last decade, authors linked to the Center for Translational Molecular Medicine (CTMM) have extended the scope of 'Early HTA' to include public decision-makers, as well as developers, as the intended audience (Ijzerman and Steuten, 2011) and to include a broad range of methods such as Clinical Trial Simulation and Option Analysis (Markiewicz et al,

2014) which may be beyond the resources of many SME developers and the expertise of some HTA analysts and could be thought of as methods of research and development or commercial activities rather than assessment. Whilst this wider scope is useful in capturing the full potential of HTA undertaken alongside the *development* of health technologies, it does result in a loss of focus on assessment activity and HTA that informs the *developer*.

In addition to the diversity of the current 'Early HTA' literature, a second difficulty with the academic literature is that it is likely to be unrepresentative of much HTA undertaken to inform developers (Grutters et al, 2019; Love-Koh, 2019). Empirical examples of work undertaken in-house, in large pharmaceutical and medical device companies, are lacking in the published literature. Moreover, most consultancy reports undertaken for small and medium-sized developers will not be published for commercial confidentiality reasons and because there is little incentive to do so. For this reason, it is reasonable to assume that the published literature available to guide an analyst is incomplete and potentially skewed towards academic publications where the resources available for analysis may exceed those available to an SME developer.

These difficulties in the literature suggest that there is a need to provide much greater clarity for developers and analysts in terms of how HTA can be used to inform development decisions in the early stages of development. This thesis proposes a three-part frameworkfor HTA to aid developers of medical technologies. The first part sets out the features of HTA to inform developers. The second proposes a generic process and the final part sets out methods of HTA which would be appropriate for analysts to use. The five case studies provide an opportunity to illustrate the frameworks of features, process and methods.

1.2 Aims and objectives

The overall aim of the thesis is to set out a framework of features, process and methods of HTA for developers of health technologies. The case studies focusing on diagnostic technologies will contribute to the limited empirical literature in this field and provide an initial, partial validation of the proposed framework.

Through an extensive literature review and informed by the case studies this thesis will address the following research questions:

1. What are the features of HTA undertaken to inform developers?

2. What is the process of HTA undertaken to inform developers?

3. What are appropriate methods of HTA to inform developers of health technologies and how do these methods link with the particular features of HTA to inform developers?

4. To what extent do the case studies demonstrate the suggested features, process and methods of HTA to inform developers of health technologies?

5. What are the wider implications of the results of the study?

6. What recommendations for policy and future research arise from the study?

1.3 Structure of the thesis

This chapter introduces the thesis.

Chapters 3-5 contain the methodological work developing the framework of features, process and methods of HTA for developers of health technologies.

Chapter 3 sets out the methodology of the literature review.

Chapter 4 addresses the first research question:

1. What are the features of HTA undertaken to inform developers?

Using insights from the literature review and the experience of the supervisors of the thesis, a list of characteristics of HTA for the developers of health technologies was developed. This chapter supports the argument that HTA to inform a commercial developer or investor has ten key features.

Chapter 5 addresses research questions two and three:

1) What is the process of HTA undertaken to inform developers?

2) What are appropriate methods of HTA to inform developers of health technologies and how do these methods link with the particular features of HTA to inform developers?

Chapter 5 completes the proposed framework for HTA to inform developers of health technologies (development-focused or DF-HTA). It sets out a generic process for DF-HTA and categorises the methods of DF-HTA identified in the literature showing how they link with the features identified in Chapter 4. .

Chapter 6 introduces the case studies, compares them to the list of features developed in Chapter 4 and suggests a process and methods appropriate for the case studies using the elements of the framework set out in Chapter 5. Chapter 6 answers research question 4.

4) To what extent do the case studies demonstrate the suggested features, process and methods of HTA to inform developers of health technologies?

Chapters 7-11 contain the case studies. Chapters 7, 8 and 9 contain case studies assessing tests of response to treatment in ovarian cancer and rheumatoid arthritis and an extension to molecular testing in colorectal cancer. Chapters 10 and 11 contain two case studies of diagnostic tests in prostate cancer and malignant melanoma.

Chapter 12 summarises the main findings of the thesis identifying the contributions made. It discusses the limitations of the work and identifies areas where further research would be beneficial as well as any policy implications, addressing research questions 5 and 6:

5) What are the wider implications of the results of the study?

6) What recommendations for policy and future research arise from the study?

1.4 What are the contributions of this thesis

The contributions of this thesis are:

1. Proposing a framework of features, process and methods of HTA to inform developers of health technologies. The list of features extends the four features first suggested by Pietzsch and Pate-Cornell (2008) to ten features. The remaining elements of the framework build on the work of the MATCH UK collaboration (Cosh et al, 2007; Vallejo-Torres et al, 2008 and Chapman, 2013) to identify a process and methods of HTA to inform developers, with a particular focus on methods suitable in resource-constrained environments.

2. Presenting five empirical examples of HTA of diagnostic technologies. There are limited examples of HTA to inform developers in the academic literature and very few of these relate to diagnostic technologies. Two generic models with potential for wider application were developed, one of which is formatted as an R Shiny app for accessibility.

2 Context

The development of health technologies is challenging. There are the challenges faced by developers of all technologies of designing a technology which is affordable, which works as planned and for which there is a demand. Over and above these generic challenges developers of health technologies need to navigate the complex international regulatory and reimbursement environment. Development of medical technologies is costly and a high proportion of both drugs and devices fail at all stages of development (Lehoux et al, 2008, DiMasi et al, 2016). Section 2.1 sets out the development process for health technologies generally with some additional detail on the development of diagnostic technologies as these are the subject of the cases studies in this thesis.

HTA is widely used to make decisions about the usage of health technologies by payers such as insurers or managed care associations or by national or regional reimbursement agencies (INAHTA, 2020). This form of HTA often takes place when the development of a technology is complete although it may also happen later in a product lifecycle. Section 0 sets out a brief introduction to HTA. It defines HTA and describes how it has traditionally been used to assess health technologies at the adoption stage (section 2.2.1). There are a number of additional challenges associated with the HTA of diagnostic devices which are set out in section 2.2.2. It has also been suggested that HTA could be used alongside the development process to inform developers and maximise the chances of success for the technology (Pietzsch and Pate-Cornell, 2008) or to allow the earlier termination of 'uneconomic projects' (Miller, 2005). A body of academic literature has built up since the late 1990s which has been termed 'early HTA' (Ijzerman et al, 2017). However, the term early HTA includes HTA addressing diverse audiences and technologies at different stages of development. Moreover, the published literature is likely to be unrepresentative of much HTA undertaken to inform developers, as both in-house analysis and commissioned consultancy reports are not often published (Grutters et al, 2019). The term, 'Early HTA' has also come to include a number of sophisticated analytical methods (Markiewicz et al, 2014) which are likely to be beyond the reach of many small and medium-sized enterprises, both in terms

of expertise and resources. Section 2.2.3 sets out a brief introduction to this literature which is further explored in Chapters 4 and 5.

2.1 Development of health technologies

2.1.1 What are health technologies?

A health technology is defined by the International Network of Agencies for Health Technology Assessment (INAHTA, 2020) as:

> 'an intervention that may be used to promote health, to prevent, diagnose or treat acute or chronic disease, or for rehabilitation. Health technologies include pharmaceuticals, devices, procedures and organizational systems used in health care.

Note that the definition is wide, including procedures and organisation as well as pharmaceuticals and devices. For the purposes of this thesis the term 'medical technology' may be considered interchangeable with health technology. The sections that follow focus on the development of drugs and medical devices. Drugs impact the patient using biochemical reactions whereas a device relies on its physicality to treat patients (Chapman, 2013). Each jurisdiction will have its own specific definitions of drugs and devices but formal definitions from the US Food and Drug Administration (FDA) (U.S. Food and Drug Administration. 2017) are given below for information. The FDA definition of a drug is a substance:

- recognised in the official US Pharmacopoeia or National Formulary
- intended for use in the diagnosis, cure, mitigation treatment or prevention of disease
- (other than food) intended to affect the structure of any function of the body
- Intended for use as a component of a medicine but not a device or a component, part or accessory of a device.

A device is defined by the Federal Food Drug and Cosmetics Act in the US (U.S. Food and Drug Administration. 2017) as:

> 'An instrument, apparatus, implement, machine, contrivance, implant or an in vitro reagent'

That meets three conditions:

- recognised in the official US Pharmacopoeia or National Formulary
- intended for use in the diagnosis of disease or other conditions or the cure, mitigation, treatment or prevention of disease
- it is intended to affect the structure or function of the body of humans.

The definition of medical devices is broad and includes large capital equipment such as imaging machines, implantable devices such as pacemakers, surgical instruments, laboratory equipment and diagnostics.

For a known disease or condition, diagnostic devices can be used for diagnosis, for screening (in an asymptomatic population), or for monitoring people who have been diagnosed and perhaps treated (Campbell and Yue, 2016). Campbell and Yue (2016) divide diagnostic devices into three broad categories: in vitro diagnostic devices (based on samples from the human body such as blood, spit, or urine), imaging systems (radiology, mammography, computed tomography), and other in vivo diagnostic devices (in or on the human body). MedtechEurope, a trade body, includes digital health technologies as a separate category of devices (MEDTECH EUROPE, 2019).

2.1.2 Market for health technologies

The market for health technologies is large and companies in this sector make an important contribution to their national economies. In 2018, the US spent 16.9% of its gross domestic product on health (OECD, 2020a, 2020b). The corresponding figure for the European Union (EU) was 8.8%. 7.2% of the health spend (equivalent to less than 1% of GDP) was spent on medical devices, 0.7% on in vitro diagnostics and the remaining 6.5% on other medical devices. 2% of the US GDP was spent on pharmaceuticals. The European medical device industry was estimated to generate revenues of USD$115 billion in 2017 (MEDTECH EUROPE, 2019) representing 27% of the global market. The US industry was estimated to generate a further 43%. The estimated number of employees in pharmaceutical companies in the EU was 750,000 and a further 675,000 were employed in medical device companies. The structure of the market differs between the pharmaceutical and medical device industries. The top ten

pharmaceutical firms have a similar market share at around 40% (ABPI, undated, accessed 31 July 2020) to the top ten medical device companies' 38.6% (statista.com, undated, a-c, accessed 31 July 2020). However, there are more small medical device companies than pharmaceutical companies. For example, in the EU in 2017 is was estimated that 27,000 companies were operating, 95% of which were SMEs (MEDTECH EUROPE, 2019) where an SME is defined as a company employing less than 250 employees and with revenue less than EUR 50 million. SMEs are unlikely to have in-house HTA expertise, hence the need for a focus on methods of HTA to address this.

2.1.3 Development of drugs and medical devices

For most medical technologies, the development process is both lengthy and costly. For drugs, the process, based on FDA approval, can take 10 to 15 years (van Norman, 2016a) including 1-3 years of pre-clinical testing, 5-10 years of clinical testing and approximately 2 years to deal with pre-market approval by regulatory authorities. A recent study estimated that the average cost of bringing a pharmaceutical to market was in excess of USD 1.3 billion (Di Masi et al, 2016). This high cost is partly driven by high attrition rates, with 10,000 potential drugs in pre-clinical testing in the United States required to bring one drug to market (Van Norman, 2016a). The pharmaceutical sector invests heavily in research and development with an estimated investment of USD 64.6 billion invested by firms in the US in 2016 and a further USD 35.9 billion invested by the government. The corresponding figures for Europe were USD 20.1 billion invested by firms and USD 11.3 billion by governments (OECD, 2019).

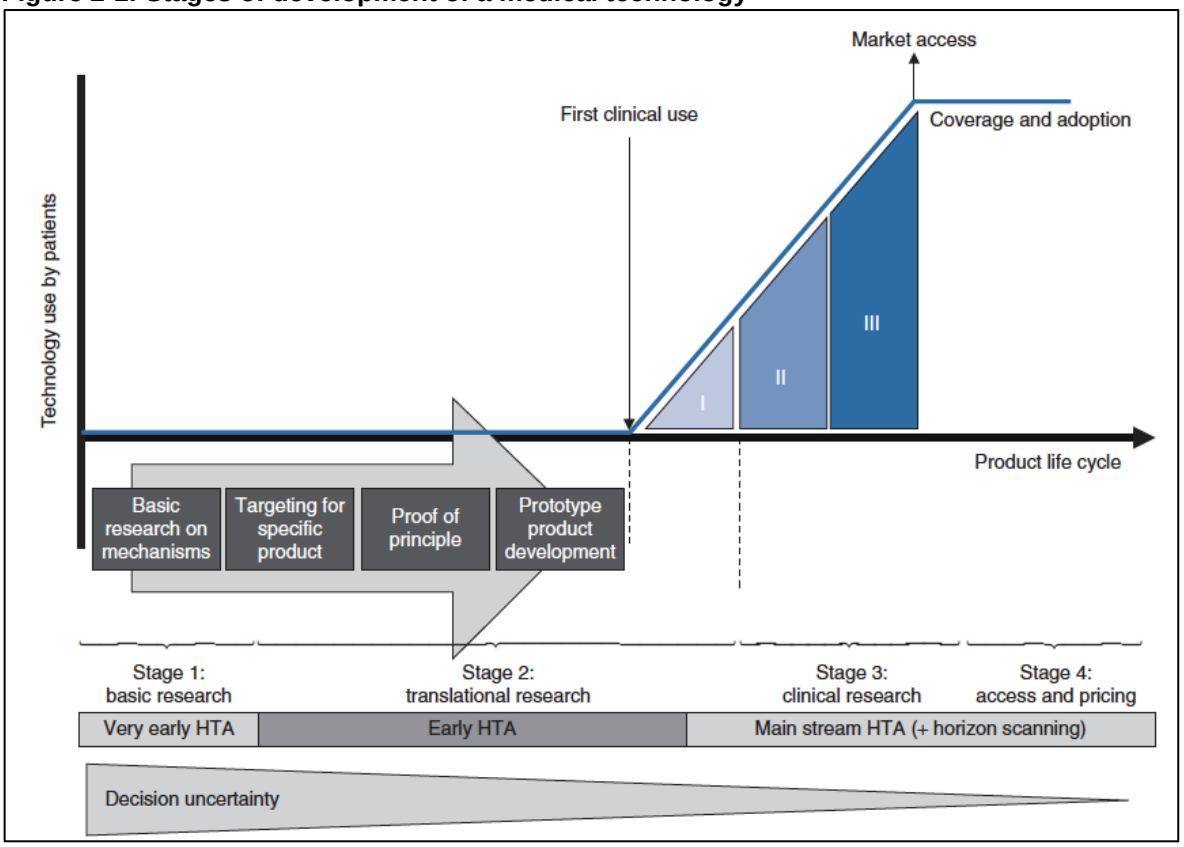
For medical devices, pathways to market access are shorter and less costly than for drugs. It has been estimated that an average time of three to seven years is required to bring a device to market (Fargen et al, 2013). In 2004, Kaplan et al found that the pre-clinical stage of development (from initiation of an idea to the construction of a prototype) takes 2 to 3 years and may cost USD 10 to 20 million. A 2010 survey of 200 medical device manufacturers found that a Class II device requiring pre-market notification in the US cost USD 31 million and a Class III device requiring pre-market assessment cost USD 94 million (MEDTECH EUROPE, 2010). It has been estimated that laboratory developed tests (LDT), which, at the time of writing, do not need FDA approval in the US, can be

developed for as little as USD 10,000 (Secretary's Advisory Committee On Genetics, Health, And Society, 2010). Diagnostic tests can be developed relatively quickly, as is evident from the recent proliferation of tests developed in response to the COVID-19 pandemic (for example, U.S. Food and Drug Administration. 2020a). COVID-19 tests have been approved for Emergency Use Authorisation in the US (and similar measures elsewhere) involving a streamlined validation process (U.S. Food and Drug Administration. 2020b).

Approximately 13,000 new patents were registered for medical devices in 2017 compared to 6,300 for pharmaceuticals. It was suggested in 2013 (Chapman, 2013) that there was less scope for development of new pharmaceuticals as there were only 218 therapeutic targets. There is some evidence that the number of therapeutic targets has increased in recent years (Santos et al, 2017; Wang, 2020) but as there are no equivalent data for devices this suggestion is difficult to support.

Figure 2-1 is an overview of the process of development of a medical technology presented by Ijzerman and Steuten in 2011. The process presented involves pre-clinical and clinical stages. The pre-clinical stage comprises basic research, targeting, proof of principal and prototype product development. The clinical stage involves testing in humans. Although this figure is representative of the process for pharmaceuticals and some medical devices there are some key differences in the development of diagnostic tests. These are highlighted in section 2.1.4.

Figure 2-1: Stages of development of a medical technology



HTA – Health Technology Assessment. Reproduced with permission from Ijzerman and Steuten, 2011.

For drugs, pre-clinical testing involves the identification of a molecule which may have potential against a therapeutic target, bench and animal testing to establish proof of principal and the development of a drug suitable to begin clinical testing. The stages of clinical testing for drugs are well established and involve Phases I to IV (van Norman, 2016a). Through these phases, drugs are tested in a sequential manner that incrementally increases patient risk. In Phase I, the drug is tested on healthy volunteers to determine the side-effect profile of the drug, the relative safety of the medication and its rate of metabolism. Phase II generally involves randomisation of patients to placebo or treatment and is designed to collect preliminary data about whether a drug is effective against a target condition. Further data will be collected about dose and safety. Phase III trials are usually head-to-head randomised controlled trials (RCTs). The new drug is set head-to-head with current standard of care in order to compare effectiveness and safety. Assuming the Phase III trial is successful the drug should be approved and Phase IV studies will be mandated to monitor the efficacy and safety of the treatment in everyday clinical practice.

For devices, Kaplan et al suggested, in 2004, that, although large companies developed successive iterations of existing devices, most new device categories were initiated by a physician or bio-engineer as a solution to a clinical problem. It is worth noting here that an alternative approach is that a technology is developed prior to identifying a specific clinical need (for example, Kluytmans et al, 2019). A preliminary prototype is built and patent process may be initiated, if appropriate. Bench testing is followed by animal testing, if appropriate, and the device enters a cycle of testing and redesign until the prototype is ready for clinical testing. One estimate of the costs of this process from 2004 was between USD 10 and 20 million (Kaplan et al, 2004), although this will vary greatly between devices. The same authors suggested that many developers seek venture-capital funding due to the scale of the costs and that devices conceived in academic medical centres would often result in license agreements to exploit the technology as academic centres often lack the capabilities to develop the device beyond early prototype (Kaplan et al, 2004).

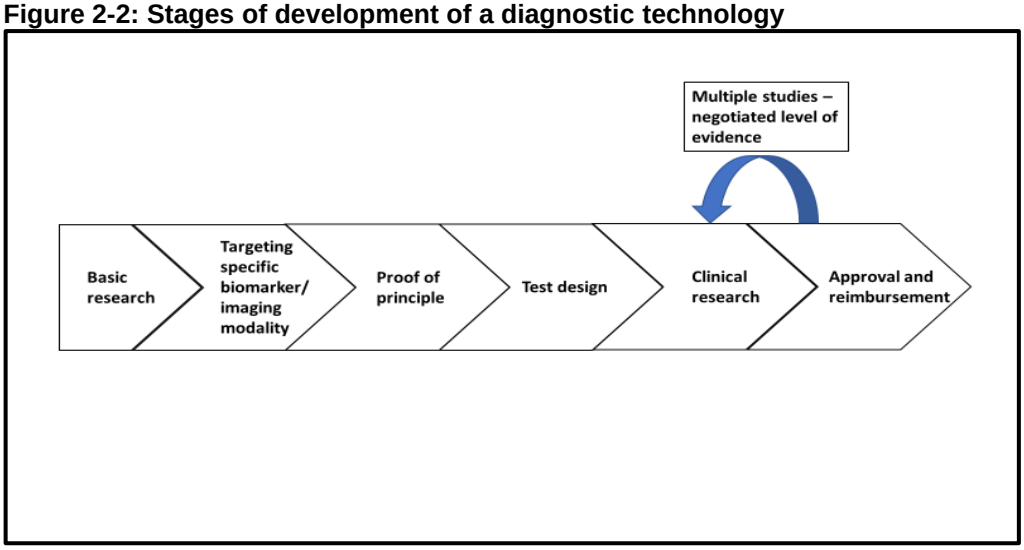
Clinical evidence requirements for market access for medical devices are determined by the nature of the device, the risk it poses to patients and criteria set by the appropriate regulator. Evidence requirements for market access for medical devices are not as stringent as the requirements for drugs and RCTs are not always required. Most high-income countries have a system of medical device regulation which follows the structure of WHO guidance originally set out by the Global Harmonisation Task Force and now embodied in The Global Model Regulatory Framework (World Health Organisation, 2017). This framework recommends definitions, classification of devices according to the risk posed to the patient and attributes of effective and efficient regulation. Marketing authorization of medical devices in the EU is based on pre-market proof of compliance with technical and safety requirements. In the EU this is called Conformité Européene (CE) which is mandatory for most categories of device. In the US, higher risk devices such as pacemakers require a pre-market application (PMA) (van Norman, 2016b). A PMA requires sufficient scientific evidence that the device is safe and effective in its intended use and will normally involve a series of studies from first-in-human use to large, multi-centre RCTs (Kaplan et al, 2004). Hwang et al (2016) found that the time to approval was longer in the

US than in the EU and that devices approved first in the EU were associated with greater risk of recall for safety issues.

Diagnostic devices have particular evidence requirements and, given the focus of the thesis, these are explained in more detail in 2.1.4.

2.1.4 Development of diagnostic technologies

Figure 2-2: Stages of development of a diagnostic technology

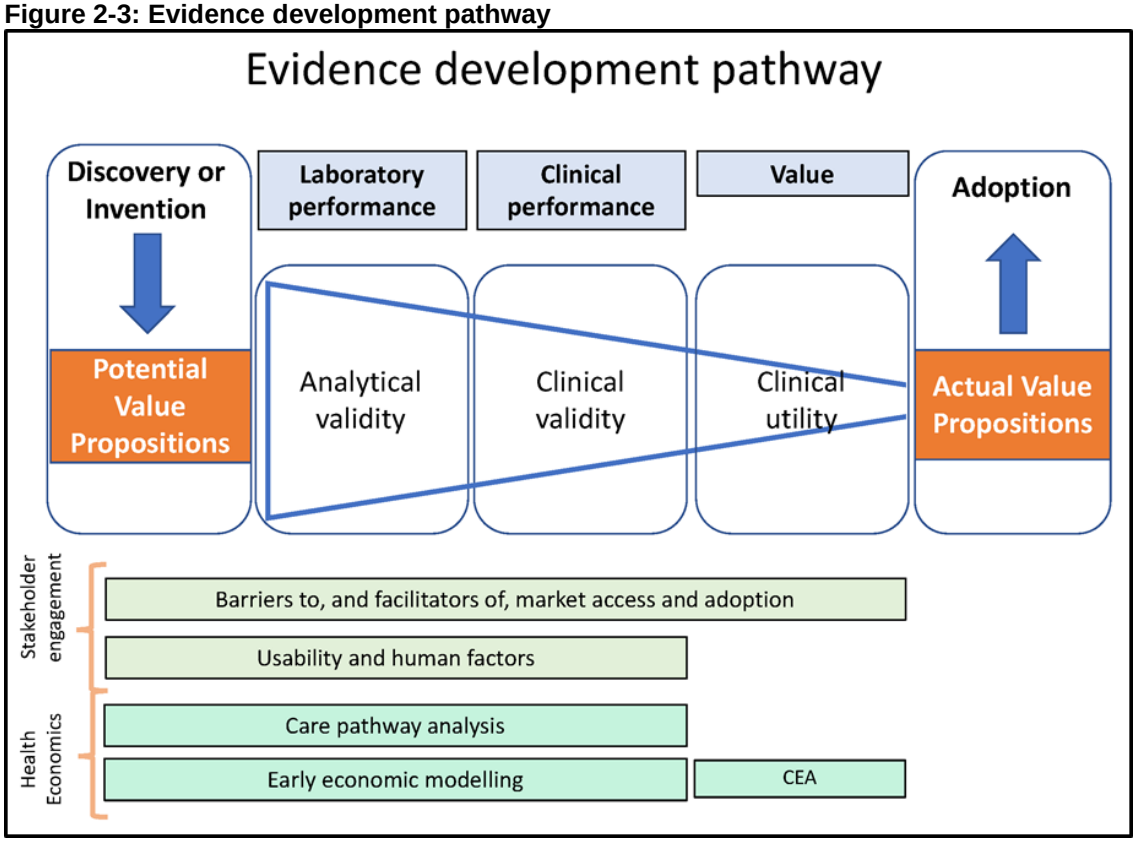


Adapted from Miquel-Cases et al, 2017, with permission

Figure 2-2 shows the stages of development of a diagnostic technology as: basic research, targeting, proof of principle, test design, clinical research and approval and reimbursement. It is based on a figure presented by Miquel-Cases et al (2017) describing the development process for predictive biomarkers. It is suitable for describing the process of development of any diagnostic technology including those using an imaging modality. Basic research involves the identification of a biomarker of interest or imaging pattern. Identification could be hypothesis-driven in that it is based on a pre-existing knowledge of the biological mechanism. Alternatively, it may be data-driven in that it involves exploratory retrospective analysis of data collected from previous studies. Following basic research, a specific biomarker/imaging modality may be targeted. Proof of principle demonstrates that the approach is feasible, allowing a prototype test to be designed ready for clinical testing. Clinical research is shown as the final step prior to approval and reimbursement. This is the step where the process for diagnostics is quite different to drugs where the Phase I to IV trials are well established. Validation of the test may include

multiple overlapping stages of evidence generation including studies of analytical validity, clinical validity and clinical utility. Different evidence requirements for different decision-makers are determined on a technology-specific basis through negotiation with regulators (Trusheim et al, 2007; Faulkner et al, 2012 and Frueh, 2013). For instance, during the current COVID-19 pandemic, the FDA have facilitated usage of diagnostic tests under Emergency Use Authorizations (EUA) setting specific requirements for analytic and clinical validation (U.S. Food and Drug Administration, 2020a). Figure 2-3 illustrates the evidence development pathway of a diagnostic. The clinical research stage from Figure 2-2 is expanded to include proof of analytical validity, clinical validity and clinical utility. Analytical validity concerns whether the test measures what it claims to measure in a laboratory setting. Clinical validity is whether the test is able to distinguish clinically meaningful populations. Clinical utility is established when the test leads to beneficial changes in treatment, outcomes and/or costs.

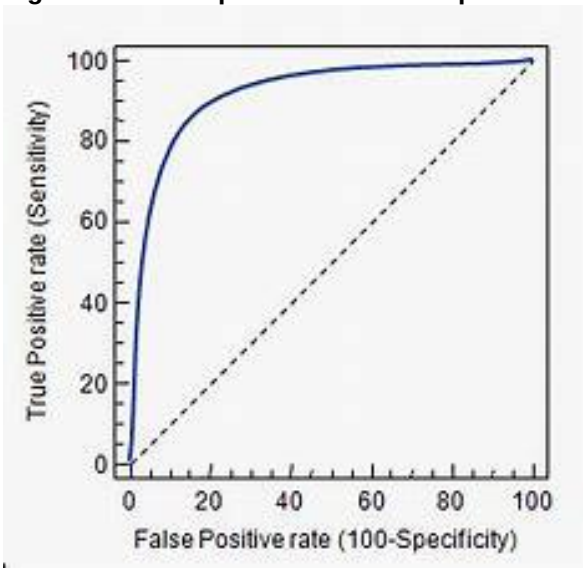
Figure 2-3: Evidence development pathway



Reproduced from Graziadio et al, 2020, no permission required. CEA – cost effectiveness analysis

Clinical studies for diagnostics are often quite different in design to those for therapeutic products (Campbell and Yue, 2016). The type of study depends on which of analytical validity, clinical validity or clinical utility the developers are trying to establish. For analytic validity they may take the form of comparing the new diagnostic test to a standard test in the same patient or the same patient's sample. The Receiver Operator Curve (ROC) is widely used in diagnostic studies where there is a reference standard (Campbell and Yue, 2016) (see Figure 2-4). The ROC curve plots the ability of the new diagnostic to identify the condition in disease-positive individuals (true positive rate or sensitivity) and to return a negative result when the disease is absent (false positive rate - specificity). The area under this curve provides a measure of test accuracy which is frequently used in diagnostic studies (Zweig and Campbell, 1993; Zhou et al, 2009). The different points (combinations of true positive rate and false positive rate) forming the ROC curve result from the movement of the cut-off threshold of the diagnostic as illustrated in Figure 2-5.

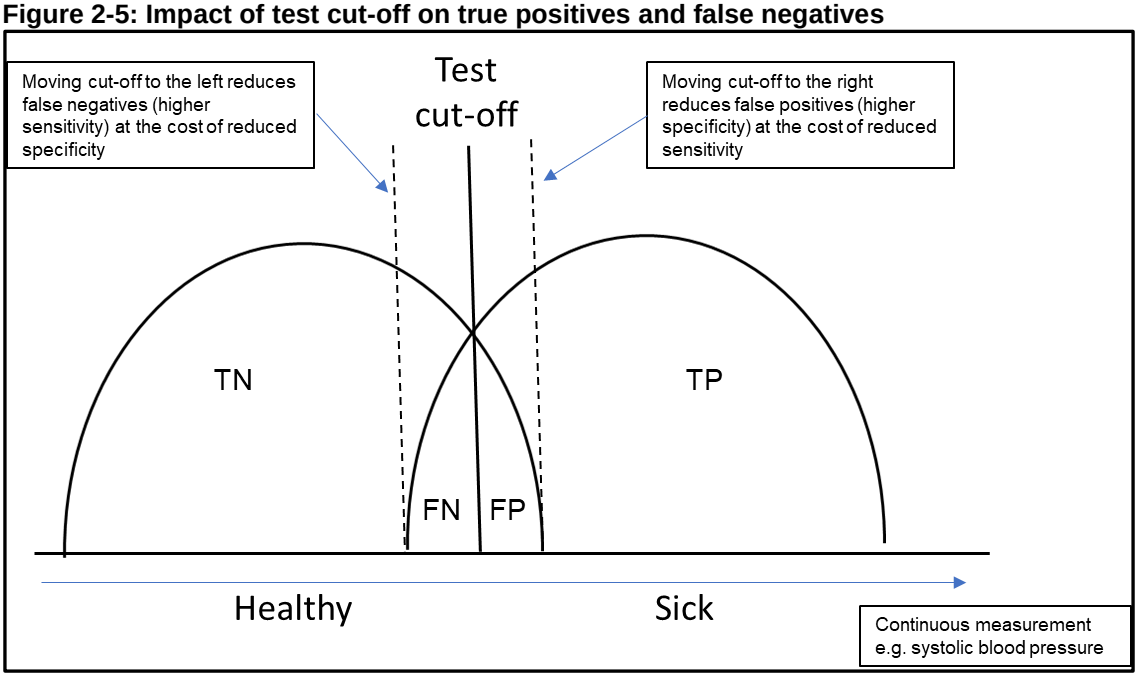
Figure 2-4: Example of a Receiver Operator Curve



Source: medcalc.org

Figure 2-5 illustrates how the movement of the cut-off threshold from left to right reduces false positives at the expense of sensitivity (true positive rate). For some uses of diagnostic tests it may be preferable to have higher true positive rate (sensitivity) to ensure that as many disease positive individuals as possible are identified. This may also result in a high false positive rate. High sensitivity may be desirable if you can give a cheap treatment, with few side effects which will prevent serious deleterious health outcomes if left untreated. High false positive rates can be an issue when the next step in the diagnostic pathway is invasive or expensive. This is the case with the Prostate Specific Antigen (PSA) test in prostate cancer (Carlsson et al, 2007). This test results in many false positives which then require further imaging and/or biopsy of the prostate in order to rule out clinically significant cancer (NICE, 2019). In other situations, a low false positive rate (high specificity) may be required. For example, where a negative result from the diagnostic test will result in the withholding of potentially curative treatment, it is important that false negatives are minimised. An example here is a prognostic test in breast cancer designed to identify women with good prognosis who may be spared chemotherapy. Effectively the optimal cut-off depends on the costs and consequences of false positives compared to false negatives. This is something which can readily be explored in HTA alongside the development process. The relationship between test threshold and sensitivity for a given ROC curve can be explored using an online interactive app from Kennis Research, 2020.

Figure 2-5: Impact of test cut-off on true positives and false negatives



FN – false negative, FP – false positive, TN – true negative, TP – true positive

2.1.5 Challenges in the development of diagnostic technologies

The case studies presented in this thesis concern diagnostic technologies, which are defined by the National Institute of Health and Care Excellence in the UK (NICE) as:

> 'all types of measurements and tests that are used to evaluate a patient's condition'.

Diagnostic is used as an umbrella term for tests which fulfil a range of roles. Table 2-1(mainly drawn from Trusheim et al, 2007) distinguishes 12 positions in the clinical pathway for diagnostic tests together with examples.

Table 2-1: Roles and examples of diagnostic tests

Position in pathway	Example
Diagnosis in asymptomatic population (screening)	Bone density test to encourage use of alendronate sodium to reduce the risk of bone fractures (Trusheim et al, 2007)
Diagnosis in symptomatic or previously screened population – rule in	Biomarker-guided surveillance for oesophageal adenocarcinoma in patients with gastrointestinal reflux (Rubenstein et al, 2005)
Diagnosis in symptomatic or previously screened population – rule out	Diagnosing and excluding ectopic pregnancy (Wedderburn et al, 2010)
Prevention – identification of individuals at risk (primary)	Identifying individuals at risk of developing Type 2 diabetes (de Graaf et al, 2015)
Prevention - identification of patients with undiagnosed disease to allow earlier treatment (secondary prevention)	Identifying individuals with undiagnosed Type 2 diabetes in order to initiate earlier treatment (de Graaf et al, 2015)
Prevention – identification of patients at risk of developing complications (tertiary prevention)	Identifying patients with Type 2 diabetes at risk of developing microvascular or macrovascular complications (de Graaf et al, 2015)
Selection of treatment or prediction of response to treatment	Testing for BCR-ABL-positive tyrosine kinase genotype to identify patients with chronic myeloid leukemia likely to respond to imatinib mesylate (Capdeville et al, 2002)
Predicting adverse events	Screening for HLA-B*5701 to prevent abacavir hypersensitivity in HIV patients (Hughes et al, 2004)
Adaptation of dose to individual patient characteristics	TPMT Genotyping for Azathioprine (Thompson et al, 2014)
Monitoring response to treatment	Bone density test to monitor response to alendronate sodium to reduce the risk of bone fractures (Trusheim et al, 2007)
Monitoring adherence to treatment	Serum concentration of anti-epileptic drugs such as phenytoin (Osterberg and Blaschke, 2005)
Predicting disease progression (prognosis)	Oncotype Dx to predict the level of risk of breast cancer progression in order to guide adjuvant chemotherapy treatment. Patients identified who are unlikely to benefit from chemotherapy can avoid the side effects and cost of treatment (NICE, 2018d)

Both licensing and reimbursement of diagnostic technologies are complex due to the difficulty in generating evidence linking health outcomes to test results. Historically, payers have not used a value-based approach in determining what they were willing to pay for a diagnostic technology. This has resulted in disincentives for developers of diagnostic technologies (Towse et al, 2013). In order to justify prices on any basis other than cost-plus, it is useful for developers to demonstrate and evidence the value added by the diagnostic technology. The value added by a diagnostic technology is highly dependent on the context in which it is to be used, and this context can vary even within a single jurisdiction. Evidence of the value generated must take account of this complexity and heterogeneity if decision-makers are to be persuaded to adopt new diagnostic technologies.

Commentators have suggested four broad challenges for developers in generating evidence about the value of a diagnostic technology. First, evidence requirements for regulation of diagnostic technologies are complex, are becoming more onerous and vary across jurisdictions (Ansari, 2013). Second, additional evidence may be required for reimbursement. Payers may require the demonstration of clinical utility before approval and this evidence is not straightforward to produce (Steyerberg et al, 2010, 2015; Doble, 2016b). Third, developers may find it hard to achieve profitable prices for diagnostic technologies due to historic cost-plus pricing and a lack of intellectual property protection (Sachs, 2015). Finally, value is difficult to define and different stakeholders may have different conceptions of value.

Evidence for regulation

The landscape of diagnostic regulation is 'constantly changing' (Ansari, 2013). The onus is on the developer to present sufficient evidence to satisfy regulatory authorities in different jurisdictions and it may require multiple studies until the diagnostic technology's safety and performance are acceptable and the evidence base demonstrates this to multiple regulators' satisfaction (Ansari, 2013). The EU recently tightened regulation of diagnostic devices. The new Medical Devices 2017/745 (MDR) and In-Vitro Diagnostic Medical Devices 2017/746 (IVDR) regulations require clinical evidence before a diagnostic can be granted a CE mark and be made available in the EU (EU, 2017). The US requires prospective clinical studies for high risk diagnostics (Sachs, 2015) whereas in the EU, clinical evidence does not necessarily take the form of prospective RCTs. In the EU, the requirements under the IVDR are for clinical evidence to support a beneficial risk-benefit balance and performance characteristics. In the US, pre-market approval requires clinical testing and evidence of efficacy and safety. Historically, some countries outside of the US and EU would have relied on certification in these jurisdictions to allow market access. This is now happening less (Ansari, 2013). More countries are introducing their own regulations and evidence requirements in large markets such as China and Japan are more stringent than those demanded by the US and Europe (Ansari, 2013). For example, in China, at least three prospective clinical studies are required before market access and in both China and Japan, studies are required

demonstrating the correlation of diagnostic test results to country gold standard. The situation in the UK is under review following the exit of the UK from the European Union with previous EU directives continuing to apply until the completion of a government review (MEDICINES AND HEALTHCARE PRODUCTS REGULATORY AGENCY, 2020).

Evidence for reimbursement

Whereas regulators require evidence of performance, clinical utility is often key for reimbursement. Clinical utility is the ability of the diagnostic technology to deliver an improvement in clinical outcomes and/or a cost saving compared to current standard of care (Steyerberg et al, 2010, 2015; Doble, 2016b). For diagnostic technologies, the ideal form of evidence of clinical utility is a prospective 'end to end' RCT where a testing strategy is compared to a non-testing strategy and all the consequences flowing from the test including treatment and health outcomes are compared over an appropriate time horizon. This kind of trial is expensive and takes a long time (Parkinson et al, 2014). As market access (i.e. regulatory approval) is possible without demonstrating clinical utility and copycat diagnostics can be produced by clinical laboratories in a relatively short time there is little incentive for developers to undertake good quality studies of clinical utility (Parkinson et al, 2014). Such evidence as there is may be poorly designed and/or reported. It has been claimed that the majority of test-treatment trials are subject to performance, ascertainment and attrition biases and sample size inadequacy (di Ruffano, 2017; Miquel-Cases et al, 2017). There is a further difficulty that the clinical utility of a diagnostic technology is highly dependent on the context in which it is used. In particular, the test may be used at different points in a clinical pathway, in combination with other diagnostic technologies with different impacts on treatment. So even if there is RCT evidence and it has good internal validity, it may not be easily generalisable (Faulkner et al, 2012). Decision-makers need to balance the requirement for sufficient evidence with the imperative to allow patients to benefit from potentially valuable technologies. In some jurisdictions, forms of evidence of clinical utility other than RCTs, such as observational data or evidence from retrospective studies may be accepted by decision-makers (Frueh, 2013; Academy of Medical Sciences, 2016).

Intellectual property protection

It has been suggested that developers of diagnostic technologies currently face disincentives derived from three sources: pricing structures; intellectual property protection and differences in regulation between laboratory developed tests and commercially developed kits (Sachs, 2015). First, reimbursement for diagnostic technologies has historically been on a cost-plus rather than a value basis and administrative codings within Medicare/Medicaid still support a cost-plus approach (Ansari, 2013; Academy of Medical Sciences, 2016). There has also been significant downward pressure on the amount reimbursed for diagnostics in the US led by a Congress committee influencing Medicaid and Medicare reimbursement rates which many private payers use as a point of reference (Sachs, 2015). Furthermore, in some jurisdictions (e.g. Spain, China, India, Brazil), companion diagnostic tests have not been reimbursed at any level and patients or pharmaceutical companies have had to pay (Ansari, 2013). Second, patent protection for diagnostic technologies in the US has been reduced. In 2010, the Gene Patents and Licensing Practices and Their Impact on Patient Access to Genetic Tests report (Secretary's Advisory Committee On Genetics, Health, And Society, 2010) explored the balance between incentives for developers of diagnostic technologies on one hand and the risks to discovery science and patient access of restrictive patents on the other. The report found that patents tended to restrict invention and patient access and that patent protection was not required to stimulate either discovery science of the development of new diagnostic tests. It has been suggested that the sentiments expressed in this report are reflected in subsequent court decisions which strongly restrict the ability of diagnostic technologies to achieve patent protection (Eisenberg, 2015). The final factor making it challenging for developers to achieve profitable prices for diagnostic tests stems from the different regulatory environment for commercially developed diagnostics and for laboratory developed tests (LDTs). Laboratories in the US are not regulated by the FDA but by the Centres for Medicare Services under the Clinical Laboratory Improvement Amendments (CLIA) regulations (CENTER FOR MEDICARE AND MEDICAID SERVICES, 2020). The tests CLIA-regulated laboratories develop are not scrutinised individually, rather it is the operations of the laboratory which are regulated. They are able to reverse-engineer commercially available kits

and provide a service-offering without obtaining FDA approval (Secretary's Advisory Committee On Genetics, Health, And Society, 2010). The FDA has undertaken a consultation but has not yet required laboratory developed tests (LDTs) to undergo the pre-market approval process required for commercially developed diagnostics (USFDA, 2018). This means that LDTs can be developed in a short period of time at greatly reduced cost (Secretary's Advisory Committee On Genetics, Health, And Society, 2010; McCormack et al, 2014) allowing other developers to benefit from the research and development effort and expenditure of the first-mover.

Differing concepts of value

The full value of a test-based technology may not be reflected in the QALY metric. Concepts such as personal utility and the use of cost-benefit analysis are contentious and under-developed in HTA (Fugel et al, 2016) meaning that the value of knowing (or not knowing), patient preferences (Oosterhoff, 2016) and spillover effects into later generations are not captured (Rogowski et al, 2015).

2.2 Health Technology Assessment

The previous sections described the process of health technology development and set out a number of key challenges faced by developers of diagnostic technologies. This thesis examines how health technology assessment could be used to help developers face those challenges by, for example, helping to shape an evidence generation strategy and the articulation and/or quantification of a value proposition. This section sets out a brief introduction to HTA. It defines HTA and describes how it has traditionally been used to inform decisions about whether or not to use, adopt or reimburse a 'finished' health technology (section 2.2.1). There is then some more detail about HTA for diagnostic devices as the case studies involve this kind of medical technology (2.2.2). The final sub-section sets out a brief introduction to the main focus of the thesis - HTA of technologies in development (2.2.3).

2.2.1 Definition and general process of HTA

The World Health Organisation definition of HTA is:

> 'the systematic evaluation of properties, effects and/or impacts of health technologies and interventions. It covers both the direct, intended consequences of technologies and interventions and their indirect, unintended consequences. The approach is used to inform policy and decision-making in health care, especially on how best to allocate limited funds to health interventions and technologies'. (World Health Organisation, 2020)

An ongoing project co-ordinated by the International Network of Agencies for Health Technology Assessment recently published a consensus definition of HTA which includes several additional and important aspects highlighted in bold below (INAHTA, 2020; O'Rourke et al, 2020).

> 'A multidisciplinary process that uses **explicit** methods to determine the value of a health technology at different points in its lifecycle. The purpose is to inform decision-making in order to promote an equitable, efficient, and high-quality health system.
>
> **Note 1:** A health technology is an intervention developed to prevent, diagnose or treat medical conditions; promote health; provide rehabilitation; or organize healthcare delivery. The intervention can be a test, device, medicine, vaccine, procedure, program or system.
>
> **Note 2:** The process is **formal, systematic and transparent**, and uses state-of-the-art methods to consider the best available evidence.
>
> **Note 3:** The dimensions of value for a health technology may be assessed by examining the intended and unintended consequences of using a health technology compared to existing alternatives. These dimensions often include clinical effectiveness, safety, costs and economic implications, ethical, social, cultural and legal issues, organisational and environmental aspects, as well as wider implications for the patient, relatives, caregivers, and the population. The overall value may vary depending on the perspective taken, the stakeholders involved, and the decision context.
>
> **Note 4:** HTA can be **applied at different points** in the lifecycle of a health technology, i.e., pre-market, during market approval, post-market, through to the disinvestment of a health technology.'

Health Technology Assessment, as a discipline, first developed in the United States when Congress requested Technology Assessment of health technologies in the mid-1970s. (Stevens et al. 2003) and the term is now internationally used. HTA draws on Evidence Based Medicine (EBM), a movement which has developed from the publication in 1972 of Archie Cochrane's 'Effectiveness and Efficiency' (Cochrane, 1972) to today's international organisation, the Cochrane Collaboration (Stevens et al, 2003). Evidence synthesis methods such as systematic review and meta-analysis are core to HTA and often form the basis for the clinical effectiveness estimates in cost-effectiveness analysis and health economic modelling.

Decisions supported by HTA include two broad categories: allocation of a set budget over a number of healthcare areas and decisions about the adoption, reimbursement or usage of individual technologies or programmes. In the first category, the decisions may involve which programmes to include in a package of Universal Health Coverage (for example, maternity care, vaccination programmes) or decisions about prioritisation within a research budget. The aim of the HTA would be to allocate the budget according to agreed criteria of effectiveness, value for money and other considerations, perhaps equity. The second category includes assessment of individual technologies, such as pharmaceuticals, to determine whether they should be adopted. Again, they would be likely to be assessed against pre-established criteria relating to evidence base, need, value for money and equity issues. Medical devices and surgical procedures could also be assessed in this way. There is also a growing interest in using HTA to determine whether a technology in current use should be excluded from reimbursement or coverage.

The components of HTA vary according to the decision-maker but HTA to inform a decision about use of a particular technology may start with the definition of a decision problem to address. Analysts may find it useful to use a structure to help them define the decision problem. PICO (Population, Intervention, Comparator, Outcome) is a general structure applied broadly in evidence-based practice such as HTA. The intervention is the technology to be assessed and the comparator is the current standard of care in that disease area. Once the decision problem has been defined the next step is synthesis of the clinical

evidence, using techniques such as systematic review and meta-analysis. Once the evidence on clinical effectiveness has been assembled and issues regarding evidence quality and generalisability addressed, cost-effectiveness can be considered. Finally, other considerations such as legality and ethics may be addressed (Eddy, 2009).

HTA informs a variety of healthcare decision-makers ranging from national healthcare providers like the National Health Service in the UK, to regional health authorities (for example in Spain and Canada) and local providers such as hospitals. Insurance companies and commercial health care providers also need to make decisions about coverage and reimbursement. HTA agencies may be established within or supported by the decision maker as with the National Institute for Health and Care Excellence (NICE) in the UK or may be external bodies such as the Institute for Clinical and Economic Review in the United States which is funded primarily by not for profit organisations (Institute for Clinical and Economic Review, 2020) and provides advice which is for guidance only. Some agencies have a strong emphasis on cost-utility analysis (for example, UK, Netherlands, Canada) and some have acknowledged a financial limit to the amount they consider acceptable to pay for each year in full health delivered by a health technology. Cost-utility analysis is the comparison of a proposed technology with the technology it aims to replace in terms of costs and health outcomes measured in Quality Adjusted Life Years (QALYs) (Drummond et al, 2015). NICE in the UK and the Canadian Agency for Drugs and Technologies in Health (CADTH) have developed reference cases which specify methods to be adopted and much work has been done on establishing best-practice through bodies such as the International Society for Pharmacoeconomics and Outcomes Research (ISPOR). There is a substantial body of work applying these methods across the range of health technologies.

2.2.2 HTA of diagnostic technologies

HTA of diagnostic technologies is broadly equivalent to HTA of other technologies but is complicated by four factors. Firstly, because diagnostic technologies provide information for the clinician and patient rather than directly influencing health outcomes (Oosterhoff et al, 2016), the clinical and cost-effectiveness of the technology depends on how the clinician and patient

use the information provided by the test (Thompson et al, 2014; Rogowski et al, 2016; Degeling et al, 2017). Despite their potential influence, behavioural aspects have only rarely been incorporated in the assessment of companion diagnostics (Degeling et al, 2017). Second, as discussed in section 2.1.4, HTA of companion diagnostics assesses the proposed test-treatment strategy or strategies incorporating the new diagnostic. This is compared to one or more current strategies used in the relevant jurisdiction. The incremental cost-effectiveness does not separate the test and the treatment in terms of value-added (Faulkner et al, 2012; Towse et al, 2013). Whilst this is not an issue in HTA undertaken to inform coverage or adoption, when prices tend to have been set, it may cause difficulties in undertaking HTA during the development process when the value-added may be used to try and guide pricing strategy. The third factor relates to the ability of the QALY to reflect the different aspects of value of a diagnostic technology. This is a contentious area. For example, it has been suggested that the QALY metric fails to capture aspects such as the 'value of knowing' (Fugel et al, 2016). This is the value to the patient of information provided by the diagnostic over and above any change in the management of the condition and is specifically recognised in Australia as an aspect of value which may need to be quantified beyond the QALY framework (Medical Services Advisory Committee, 2017). The final factor concerns the difficulties in generating evidence of clinical and cost effectiveness, which were discussed above in section 2.1.4. These difficulties include additional complexity due to considerations of test performance (Payne et al, 2013), multiple test combinations (Degeling et al, 2017) and how to establish appropriate clinical pathways for both intervention arm and comparators (Abel et al, 2019; Graziadio et al, 2020).

2.2.3 HTA of technologies in development

HTA is most often used around the time of market access to inform public payers about whether and how a technology should be adopted and used. It has also been recognised that there may be benefits to undertaking HTA at different points alongside the development process. Annemans et al (2000), for example, suggested that pharmacoeconomic analysis could be undertaken when a promising drug was undergoing the earliest clinical trials. Although there would be gaps in the evidence, the evaluation would allow developers to consider at an

early stage whether the drug was likely to be cost effective in a given jurisdiction and help shape evidence generation strategy. It was recognised that multiple models could consider different indications, jurisdictions and dosages and help with the direction of development of the drug (Annemans et al, 2000). Authors associated with the Multi-disciplinary Assessment of TeCHnology (MATCH) programme, in the 2000s, extended this concept to the use of HTA methods to inform SMEs developing medical devices (for example: Cosh et al, 2007; Craven et al, 2009; Chapman, 2013).

It was Pietzsch and Paté-Cornell (2008) who first used the term 'early HTA'. They also set out four characteristics distinguishing 'early HTA' from 'classical HTA' (as they termed HTA undertaken to inform payers at the time of market access – see section 2.2.1). The four elements, set out in Table 2-2 were: the aim of the HTA, the decision support provided, the available evidence and the influence of the HTA on the technology performance. Whilst the aim of classical HTA is to assess the 'safety, effectiveness and cost-effectiveness profiles of a new technology', the aim of early HTA is to assess the '(likely) safety, effectiveness and cost-effectiveness'. Whilst classical HTA provides decision support for regulators, payers and patients about market clearance, payment and usage of a technology, early HTA provides decision support for manufacturers and investors about the design and management of a technology as well as regulatory and reimbursement strategy. Available evidence for early HTA is bench and animal testing, early clinical experience and evidence from previous generations of the technology compared to evidence from clinical studies performed with new technology for classical HTA. Finally, early HTA has the ability to influence the future clinical performance of a new technology whereas classical HTA has limited or no influence on performance. In Pietzsch and Paté-Cornell's case study (2008), they suggested that the complex model built could potentially influence the design of the technology by giving feedback to the development team about the performance levels, and in this case the training standards, which would be required to ensure the technology met the minimum profit requirements of the developer.

Table 2-2: Similarities and differences between 'Classical HTA' and 'Early HTA'

	Classical HTA	Early HTA
Aim	Assess safety, effectiveness and cost-effectiveness profiles of a new technology	Assess safety, effectiveness and cost-effectiveness profiles of a new technology
Decision support	Decision support for regulators, payers and patients about market clearance, payment and usage of a technology	Decision support for manufacturers and investors about the design and management of a technology, as well as regulatory and reimbursement strategy
Available evidence	Usually evidence from clinical studies performed with new technology'	'Evidence from early bench and animal testing, early clinical experience and evidence from previous generations of the technology
Influence on technology performance	Limited or no influence on clinical performance of a new technology	Potentially significant influence on (future) clinical performance of a new technology

Reproduced from Pietzsch & Paté-Cornell, 2008, with permission. HTA – Health Technology Assessment

In the last decade, authors linked to the Center for Translational Molecular Medicine (CTMM) have extended the scope of 'Early HTA' to include public decision-makers, as well as developers, as the intended audience (Ijzerman and Steuten, 2011) and to include a broad range of methods (Markiewicz et al, 2014). Whilst this wider scope is useful in capturing the full potential of HTA undertaken alongside the *development* of health technologies, it does result in a loss of focus on HTA to inform the *developer*. The 'Early HTA' literature, as captured by Ijzerman et al (2017), is diverse in audience, timing, methods and terminology. As a result, it does not serve as a clear guide to the analyst on how to approach HTA to inform developers of medical technologies. In this thesis, the emphasis is on how HTA can inform the developer of a health technology. The key stakeholder is the developer. It is, therefore, necessary to 'unpick' the body of work badged as 'Early HTA' and clarify which aspects of are relevant to this key stakeholder.

3 Methodology of literature review

3.1 Purpose of the literature review

A literature review was undertaken to inform the first three research questions:

1. What are the features of HTA undertaken to inform developers?

2. What is the process of HTA undertaken to inform developers?

3. What are appropriate methods of HTA to inform developers of health technologies and how do these methods link with the particular features of HTA to inform developers?

3.2 Choice of literature review methodology

3.2.1 Previous and alternative approaches

Similar questions to the research questions above were addressed in four recent review articles that all used a database searching approach in their literature reviews (Hartz and John, 2008; Mikudina and Redekop, 2013; Markiewicz et al, 2014; Ijzerman et al, 2017). They used a range of search terms to describe the broad themes of 1) a health technology 2) during development 3) assessment/evaluation. Ijzerman et al (2017) added a category of known methods of early HTA such as headroom and multi-criteria. In each case the database searches were supplemented either by searches of key references of identified articles (Hartz and John, 2008; Markiewicz et al, 2014; Mikudina and Redekop, 2013; Ijzerman et al, 2017) and hand searches of key journals (Hartz and John, 2008). Hartz and John (2008) and Mikudina and Redekop (2013) also searched Google Scholar and the websites of Horizon Scanning agencies. Quality criteria were not applied by any of the authors to exclude studies because of the lack of a suitable instrument (Markiewicz et al, 2014) or because they were interested in identifying all studies rather than utilising the results of the studies.

Study authors of these review articles raised two main issues with the use of a literature search based on database searches in this area. The first issue was the sensitivity of the search due to the lack of clear definition of early HTA in Medical Subject Headings (MeSH) or clear key words (Markiewicz et al, 2014; Ijzerman et al, 2017). The second issue (Hartz and John, 2008; Markiewicz et al, 2014) was publication bias as an issue as companies would be unlikely to publish reasons for discontinuation of a project and much information would be confidential for commercial reasons.

Further generic issues with systematic database literature searches are raised by Boell and Cecek-Kecmanovic (2010). In particular, they raise the concern that, as research questions are necessarily fixed before the literature review starts, the researcher may be inhibited from pursuing further literature and the approach discourages learning from adjacent areas. They claim that because databases are limited in their journal coverage and search terms are indeterminate, all relevant studies are unlikely to be retrieved. Moreover, elaborate search strategies with high sensitivity come at the cost of low precision with only a small percentage of all retrieved results being relevant.

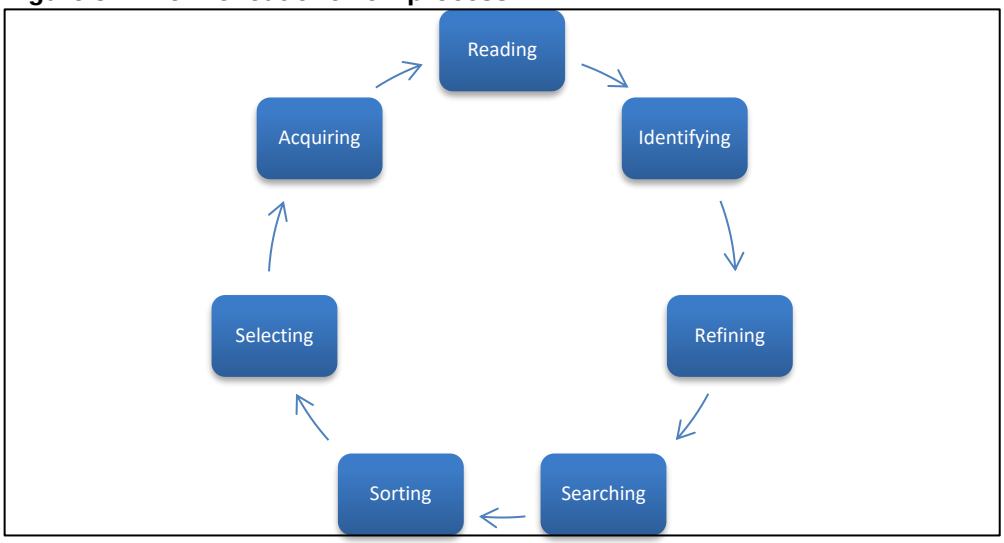
Two approaches to literature review may address some of these drawbacks, pearl-growing and hermeneutic literature review. Both of these approaches have previously been applied in health research. By way of example, Claxton et al (2015) applied a pearl-growing approach in their work to estimate the cost-effectiveness threshold for health care in the United Kingdom as the aim was to provide an overview of a wide-ranging area and any database search using search terms ran the risk of missing important parts of the literature as well as generating high numbers of papers to review. They describe pearl-growing as

> ‘the use of existing collections of studies to identify additional relevant parts of the literature. The approach uses a pool of ‘initial pearls’ to grow the literature both through references and citations until all relevant papers have been discovered. This approach, therefore, relies on the expertise of the authors of the existing literature to populate the pool of studies rather than the searcher’s potentially limited knowledge’ (Claxton et al, 2015).

Greenhalgh and Shaw (2017) applied the hermeneutic approach in exploring the potential impact of telehealth technology in a population of patients with heart

failure as their aim was to provide a: 'scholarly synthesis of the key questions, theoretical perspectives and empirical findings on the topic'. Hermeneutics is concerned with the process of creative interpretive understanding (Boell and Cecek-Kecmanovic, 2010). A hermeneutic review involves a circular and open-ended process where understanding from reading allows a refinement of the research question and strategy and allows researchers to broaden the search when investigating general relationships or narrow it when a comprehensive survey of a particular aspect is desired (Boell and Cecek-Kecmanovic, 2010). Iterations of the process continue until saturation of understanding is reached (Boell and Cecek-Kecmanovic, 2010). The hermeneutic review process is illustrated in Figure 3-1 (Boell and Cecek-Kecmanovic, 2010).

Figure 3-1: Hermeneutic review process



Adapted with permission from Boell and Cecek-Kecmanovic, 2010.

3.2.2 Literature review method adopted in this study

From previous authors' experience it would be unlikely that a database search with appropriate sensitivity and specificity would be achieved. An alternative approach was taken which combined pearl-growing (i.e. following references and citations from key articles in the field) with the hermeneutic approach.

Ijzerman and Steuten's 2011 paper was used as the 'pearl' so formed the basis of the search. It was selected as it was a comprehensive review article in the field of interest which provided an introduction to the intellectual concepts of the area and the structure of those concepts (Boell and Cecek-Kecmanovic, 2010). The references and citations of this work on Google Scholar were taken

as the first source for the review. Citations on Google Scholar were compared to other databases (Web of Science and Scopus) and Google Scholar was found to be more comprehensive.

The aim was to retrieve literature which informed an understanding of early health technology assessment (as much early HTA is intended to inform developers) and included both methods papers and applied studies separately and in combination. Inclusion criteria for applied studies were that a study needed to include an application of early HTA. Methodological papers were included where they introduced a new method or discussed a known method of early HTA. A judgement was then made whether the early HTA paper was also intended to inform developers. A paper was included if it was explicitly aimed at the developer of the technology or if the decisions it informed concerned aspects of a technology in development other than its cost-effectiveness in a given jurisdiction at a fixed price. No quality criteria for inclusion were applied as the aim was to increase understanding of methods in use rather than rely on any conclusions reached in the studies. Abstracts of citations were reviewed and relevant articles were selected for full review. Where sufficient understanding of a method and the context of its use had been achieved no further examples were sought. Continuing the pearl-growing approach the references and citations of the literature acquired in previous steps were reviewed until saturation had been reached. This point is reached when additional publications make only a 'marginal contribution to further understanding' (Boell and Cecek-Kecmanovic, 2010). The main literature searches were conducted in October 2017 (see Appendices 1 and 2) with the citations of all studies listed in appendices 3 and 4 rechecked in February 2019.

3.2.3 Results of the literature search

See Appendix 1 for a flowchart of the pearl-growing literature review with papers retrieved at each stage in each category of interest. A total of 152 papers were identified as set out in Table 3-1. Lists of the papers identified in each category are provided in Appendix 2.

Table 3-1: Results of literature search

	Methods papers	Applied papers	Methods and applied papers	Total
Early HTA	56	61	35	152
Of which HTA to inform developers	43	25	20	88

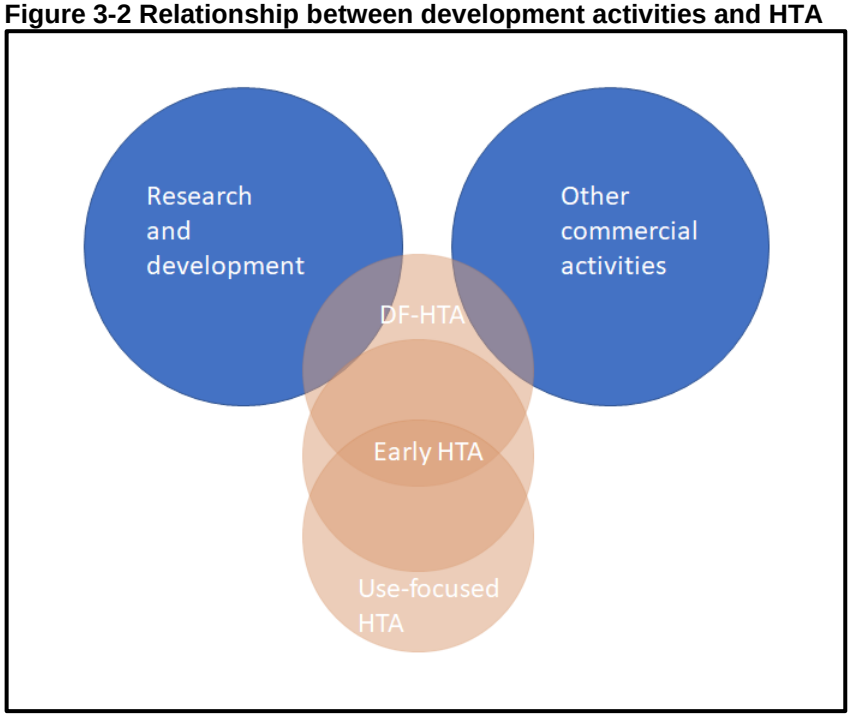
Chapters 4 and 5 discuss the papers retrieved in the literature search in order to inform the development of a framework of features, process and methods of HTA to inform developers. Early HTA papers are retained in the dataset as there is much overlap between some features, process and methods of early HTA and HTA to inform developers. Moreover, two of the case studies could be classed as early HTA rather than HTA to inform developers given that they were primarily concerned with a reimbursement perspective. It is thus, informative to have both sets of papers.

3.2.4 Limitations of the literature review

The literature search may not have identified all relevant methods and applications of HTA to inform developers. The aim of the search was not to identify all applied papers but to gain an overview of methods in use. Two recent systematic reviews (Ijzerman et al, 2017 and Markiewicz et al, 2014) were included in the review, both of which used broad database search approaches. It is, therefore, likely that methods represented in the literature have been covered. However, the academic literature is likely to contain only a small proportion of HTA undertaken to inform developers. This is because work undertaken to inform developers is generally both commercial in confidence and may also not be suitable for publication due to the use of what may be perceived as less robust methods at earlier stages of development. Published work appears to include much work supported by translational research bodies where more resources may be available for analysis.

A difficult feature of the literature review involved setting the scope of what constituted HTA, early HTA and HTA to inform developers and how HTA fitted with and could be distinguished from research and development activities and commercial activities in medical technology firms. This is important as early HTA is not a well-defined term (Ijzerman and Steuten, 2011) and many of the methods previously identified in the literature (e.g. Markiewicz et al, 2014) as early HTA would appear to be methods of research and development or commercial activities rather than assessment activities. This makes it difficult for analysts to determine what methods may be appropriate in undertaking HTA to inform developers. **Error! Reference source not found.** puts forward a suggestion of the relationship between these various activities.

Figure 3-2 Relationship between development activities and HTA



DF-HTA – development-focused HTA, HTA – health technology assessment. This figure illustrates the relationship between the research and development process, other commercial activities and HTA. It should be noted that these activities are all interlinked and sharp distinctions have been shown for didactic purposes only. There is an overlap between research and development and DF-HTA as, for example, some stakeholder consultation exercises will not only allow assessment of the technology but will also contribute to its design/usability. There is overlap between other commercial activities and DF-HTA as an activity such as strategic planning, for example, may contribute to the identification of scenarios to model in DF-HTA as well as assisting the developers' marketing strategy. There is overlap between the various kinds of HTA. Much DF-HTA is also Early HTA as there is little evidence specific to the technology. Some DF-HTA is also useful to users so there is overlap between DF-HTA and use-focused HTA.

One category of studies identified as early HTA (Ijzerman and Steuten, 2011, Markiewicz et al, 2014) but generally undertaken to inform reimbursement

agencies was horizon scanning activities. The approach taken with horizon scanning was to exclude it from the literature review.

A greater difficulty was in setting a boundary between HTA activities and research and development activities such as portfolio management and technical/engineering product development activities. Some methods identified as early HTA in methodological articles, such as clinical trial simulation (Hartz and John, 2008; Ijzerman and Steuten, 2011; Mikudina and Redekop, 2014) and real options analysis (Bartelmes et al, 2009; Ijzerman and Steuten, 2011; Markiewicz et al, 2014) appeared to fall outside the boundary of HTA. The approach taken with both these methods was to exclude both clinical trial simulation and real options analysis and similar methods.

4 Developing a framework of HTA to inform developers: part one - features

4.1 Introduction

Chapter 4 addresses the first research question:

1. What are the features of HTA undertaken to inform developers?

An adapted form of this chapter was published in 2020 in the International Journal for Technology Assessment in Health Care.

HTA conducted to inform developers of health technologies is typically characterised in the academic literature as "early" health technology assessment. The label "development-focused HTA" (DF-HTA) is preferable as is it explicitly describes the purpose of the analysis, rather than alluding to just one, arguably not defining, characteristic. DF-HTA has a number of distinct features when compared to HTA conducted to inform reimbursement and usage decisions (use-focused HTA). In order to conduct effective DF-HTA, it is important that analysts, who are often more familiar with use-focused HTA, are made aware of these differences. This is particularly important as DF-HTA analyses conducted or commissioned by commercial technology developers and in-house analyses undertaken in pharmaceutical and large medical device companies are typically not published due to a desire to maintain confidentiality and lack of incentive to publish (Grutters et al, 2019). Assessments of medical devices developed by SMEs or academic groups may be published if some public funding has been provided. Consequently, the studies which are published are likely to be a biased sample of the work undertaken and may not be particularly useful as a reference source for HTA analysts new to working with developers.

There is some useful methodological content in the published literature. For example, the MATCH collaboration in the UK aimed to support companies in the UK healthcare technology sector to assess the value of medical devices from concept through to mature product. The collaboration extended the concept of

iterative economic evaluation described in the late 1990s and 2000 (Sculpher et al, 1997; Annemans et al, 2000) to develop methods and tools (for example, Cosh et al, 2007; McAteer et al, 2007; Vallejo-Torres et al, 2008; Girling et al, 2010; Chapman, 2013) for SMEs. The SMEs were often working in resource-constrained environments and had little in-house knowledge or experience of HTA (Craven et al, 2012). Other groups, particularly translational research bodies such as ProHTA (Kolominsky-Rabas et al, 2015) and the Center for Translational Molecular Medicine (Steuten, 2016), built upon the MATCH work and published further methodological and applied papers (for example, Ijzerman and Steuten, 2011; Markiewicz et al, 2014; de Graaf et al, 2018). This form of HTA, initially described as 'supply-side' HTA by McAteer et al of the MATCH collaboration (McAteer et al, 2008), has more recently been described as 'early HTA' (Pietzsch and Paté-Cornell, 2008; Ijzerman and Steuten, 2011).

The term "early HTA" as a term to describe HTA undertaken to inform developers is somewhat vague and unhelpful. It might be taken to imply that methods and approaches for development-focused HTA are essentially similar to those of the more commonly reported use-focused HTA, only undertaken at an earlier point in time. This problem is compounded by the fact that many early health economic modelling studies in the published literature take the normative structure of use-focused HTA and apply it at an earlier point in time. However, there are fundamental differences between development- and use-focused HTA that arise as a consequence of the differences in the target audience and the decisions that the analysis is intended to inform and these are more important than the timing of the analysis.

This chapter aimed to produce a characterisation of DF-HTA that is useful to analysts new to working in this field as the first element of a framework of DF-HTA. The characterisation, in the form of a list of features and accompanying questions, was intended as an aide-memoire for analysts more familiar with use-focused HTA. The list and questions could be used at the outset of a project to prompt reflection on the nature of the assessment and to help structure discussions with developers. It could also be used at the reporting stage of a project, whether published or not, to ensure transparency. The suggested list of features may also form the starting point for a debate in the wider academic

community about the nature of HTA undertaken to inform developers of health technologies.

4.2 Features of DF-HTA

4.2.1 Methods

An iterative process similar to framework analysis was undertaken (Spencer and Ritchie, 2002; Oliver et al, 2008) to develop the list of features. This was informed by selected papers from the literature review described in Chapter 3. An initial list of features was developed, informed by two explicit frameworks identified in the literature (Ijzerman and Steuten, 2011; Pietszch and Paté-Cornell, 2008; see Appendix 3) and the prior experience of Professors Briggs and Hawkins, supervisors of this PhD thesis. The initial list was refined and expanded using an iterative process informed by the content of articles identified as being particularly informative (see Appendix 4). These papers were selected as they discussed DF-HTA in sufficient detail to allow additional features to be identified or provide new insight on previously identified features. The list of features was compared against the identified applied papers (see Appendix 5).

4.2.2 Results

4.2.2.1 Features of DF-HTA

Two studies set out explicit frameworks distinguishing between different perspectives of HTA and a further 11 discussed the distinction sufficiently for an implicit structure to be discerned. These studies are set out in Appendices 3 and 4. The two studies with explicit frameworks (Pietzsch and Pate-Cornell, 2008 and Ijzerman and Steuten, 2011) set out four features: perspective (user), timing, decisions informed and available evidence. A further four features were added after examination of the papers including implicit structures. These were: underlying user objective; decision space; business model; and resources for analysis. The final two features were added following discussion with PhD supervisors which were stance of analysis and burden of proof. The ten features are presented in Table 4-1. The identified features are not separate and independent but intimately linked. For example, evidence is lacking because of

the timing of the assessment. However, they are each worthy of explicit consideration.

Table 4-1: Features of development-focused HTA

Feature of DF-HTA	Description
Target audience	Technology developers (both academic and commercial) and investors (both commercial and public sector)
Underlying user objective	Commercial developers and investors maximise long-term financial return on investment Public funders and non-commercial developers maximise societal return on investment, health or other goal, such as employment levels or financial growth
Decisions HTA designed to inform	Broad range including: Pre-clinical/preliminary market assessments First estimations of pricing/reimbursement scenarios Go/no go decisions Technology design Trial design/evidence generation strategy Research prioritisation
Decision space	Wide including multiple: Jurisdictions Indications Comparators Funders User groups Thresholds (test cut-off) Levels of test performance Positions in pathway Technologies (if a portfolio of research projects is being assessed)
Available evidence	Clinical studies tend to be small such that uncertainty is high Evidence specific to technology scarce early in the development process. Alternative methods of estimating parameters include: Expert opinion Evidence on comparators or previous generations of a technology Bench or animal studies Output from pharmacodynamic models Evidence required about usability and clinical pathways
Timing	Repeated on an iterative basis Pre and during development
Business model	Fluid -not yet defined Various business models available including reimbursement-based models, direct marketing to patients, clinicians or health-care organisations
Resources for analysis	Often constrained at early stages due to conflicting demands on resources Less resource-intensive methods to establish and begin to quantify value proposition
Stance of analysis	Positive Which jurisdiction, position in pathway maximises return for developers?
Burden of proof	"Consumer-specific" methods and evidence credible to the development team Limitations made transparent

DF-HTA – development-focused HTA, HTA – Health Technology Assessment

1. Target audience

The target audience for DF-HTA includes both the developers of technologies and the sponsors or funders of the development. These may include both commercial and academic institutions as well as private and public sectors funders (Pietzsch and Paté-Cornell, 2008). Hereafter, the general term "developers" will be used. The target audience of developers is the defining feature of this form of HTA. DF-HTA differs from other forms of HTA because of the requirements of its target audience. In published studies, the target audience is often not explicitly defined (e.g. Hjelmgren et al, 2006; van Nimwegen et al, 2017; Vilsbøll et al, 2018). In some cases, the analysts appear to adopt the perspective of a payer even when the HTA is undertaken to inform the developer. For example, Latimer et al (2011) undertook an economic evaluation to inform developers about the feasibility of designing a collar for use by patients with motor neurone disease which would be cost-effective from the perspective of UK NHS. Such an analysis fails to explicitly recognise that the technology might be marketable in multiple markets that apply differing criteria to determine reimbursement.

2. Underlying user objective

The primary objective for a commercial sector developer or investor is to maximise long-term financial return on investment ((Ijzerman and Steuten, 2011; Hartz and John, 2008; Girling et al, 2015; Buisman et al, 2016b)). Other social objectives or motivations are typically subservient to this objective. The primary objective of public sector developers (for example, academic developers funded by public bodies) is to maximise the societal return on investment. Societal return includes consideration of direct financial returns on development, industrial growth or employment and improvements in societal health (e.g. Innovate UK's funding streams (Innovate UK, 2018)). In order to maximise long-term financial return on investment, developers and investors need to consider the measures of value for money that payers use in their coverage decisions. Thus, the underlying objective of the payer is relevant for developers to inform pre-clinical, preliminary market assessments and first estimations of pricing and reimbursement scenarios. Although explicit thresholds, such as the £20,000 to £30,000 per quality adjusted life year in the

United Kingdom (NICE, 2013), are often used in DF-HTA to make a first estimate of the maximum price achievable for a technology in order for it to be considered cost-effective (McAteer et al, 2007; Chapman, 2013; Markiewicz et al, 2014) it is important for DF-HTA analysts to recognise that a range of approaches are used by different payers. The underlying decision rules used by decision-makers should reflect their objectives. Analysts undertaking DF-HTA should acknowledge that commercial developers and investors will use, either implicitly or explicitly, a decision rule based on the expected net present value of an investment. This means that they will be interested in the expected revenues to be generated across relevant markets and the expected costs associated with delivering these sales as well as the timescales over which the revenues and costs occur. In principle, the net present value of these potential inflows and outflows should be calculated using a discount rate which takes into account the company's cost of borrowing reflecting the perceived risk of the project. In practice, crude measures of the opportunity costs of a particular investment are likely to be used. Public funders and non-commercial developers may base a decision to continue the development on a formal net value of information analysis based on the acceptable cost-effectiveness threshold (which itself should represent the opportunity cost of healthcare expenditure). More informal analyses may simply try to estimate the likelihood or potential that a technology will be regarded as cost-effective. In this situation, the decision rule used is to continue the development if the technology is likely to meet the appropriate thresholds for cost-effectiveness in the relevant jurisdiction. Commercial developers and investors may also be interested in the outcome of this analysis as it would provide some indication of the likelihood of sales in the relevant jurisdiction and provide guidance as to acceptable pricing. It should be recognised that explicit thresholds are not used by all payers, they vary between jurisdictions, they are not the only determinant of reimbursement, and they are subject to change. Commercial developers and investors may also take into account other aspects of value not typically included in formal cost-effectiveness analyses, such as patient convenience or comfort and the value of knowing a diagnosis for patients and their families, if these are likely to influence usage (van Nimwegen et al, 2017; Rogowski et al, 2016).

3. Decisions HTA designed to inform

DF-HTA potentially informs a wide range of decisions and considerations including: preliminary market assessment; estimation of pricing; review of reimbursement scenarios; individual go/no go decisions; technology design; evidence generation strategy including study design; and research and development portfolio prioritisation (Pietzsch and Paté-Cornell, 2008; Hartz and John, 2008). As DF-HTA is undertaken before the development process concludes, developers can respond to the assessment by changing the design of the technology, its target indication(s) and position in the clinical pathway (Pietzsch and Paté-Cornell, 2008). The assessment process itself may highlight gaps in the evidence for the new technology which can drive the evidence generation strategy at the next phase of development. This can also facilitate discussions with regulators or reimbursement agencies that increasingly offer to engage with developers during the development process. If assessment is undertaken simultaneously for a number of technologies, the results can be used to identify the most promising technologies facilitating the prioritisation of research effort and expenditure. For example, de Graaf et al (2018) assessed the potential of biomarker tests in four roles in the prevention of Type 2 Diabetes Mellitus to prioritise research effort and expenditure within a translational research organisation.

4. Decision space

Decision space means the range of different ways and places in which a technology may be used, for example, clinical indication, target population, and placement in the treatment pathway. In DF-HTA, the decision space is often wide and poorly defined. As DF-HTA is generally undertaken prior to licensing, the potential indications and positions in the clinical pathway are not yet constrained by licensing restrictions and multiple options may need to be assessed (Annemans et al, 2000). Other aspects of decision space include multiple versions of the technology (including optimisation of test characteristics for diagnostics) (Buisman et al, 2016b), patient populations (Hartz and John, 2008; Buisman et al, 2016b), jurisdictions, comparators, dosages, modes of delivery, pricing structures (Annemans et al, 2000) and diffusion scenarios (Rogowski et al, 2016). Furthermore, these may vary across different potential

markets. If a portfolio is being assessed decision space will include multiple technologies or potential research projects.

5. Available evidence

In DF-HTA, evidence specific to the technology is typically scarce early in the development process. As direct evidence of clinical effectiveness is lacking there is more reliance on elicited expert opinion (Cosh et al, 2007, Vallejo-Torres et al, 2008), evidence relating to comparator technologies (Vallejo-Torres et al, 2008), bench or animal studies, previous generations of a technology and extrapolations from pharmacodynamic models (Annemans et al, 2000). Where direct clinical evidence is available, studies are often small so that uncertainty around any estimates is high. Methods of expert elicitation have been developed to improve the reliability of experts' estimates of plausible ranges. Evidence may also be required about usability or the impact of a technology on clinical pathways (Abel et al, 2019). Qualitative methods (Davey et al, 2011; Kluytmans et al, 2019) and multi-criteria decision analysis (Hummel et al, 2000a) have been used to address this need. Shortage of evidence is not unique to DF-HTA, as uncertainty is inherent in all HTA. However, the shortage is likely to be more pronounced earlier in a development process.

6. Timing

DF-HTA is an ongoing activity facilitating a continuous discussion around the technology development process rather than a discrete event with a specific output. The majority of DF-HTA will be undertaken before a technology is approved by a regulatory body. The starting point for the DF-HTA may be the identification of a clinical need preceding the product development process (Yock et al, 2015). In this case, the DF-HTA would assess the potential for the technology proposed. An example of this approach is provided by Brandes et al (2015), who assessed a hypothetical vascular closure device and found only a single sub-group where the technology had potential. Alternatively, the starting point for DF-HTA may be the evaluation of a technology already in development. Kluytmans et al (2019) evaluated a surgical device at prototype stage and found that there was little potential for the device in meniscus surgery, which was the developers' suggested indication. DF-HTA is particularly suited to an iterative

approach with discussions with developers continuing alongside the development process and analysis undertaken prior to significant investments, such as Phase II or Phase III trials for pharmaceuticals (Vallejo-Torres et al, 2008; Girling et al, 2010). Vallejo-Torres et al (2011), presented an iterative economic evaluation of absorbable pins for hallux valgus at three different stages of development. The authors used retrospective data for this analysis to recreate the dynamic process of DF-HTA occurring in real-time alongside the development process. It should also be noted that use-focused HTA may also use an iterative approach (rather than the discrete event with a specific output described above), as products are arriving to market with greater levels of uncertainty.

7. Business model

In this context the term 'business model' broadly refers to how a technology and the customer are brought together, which determines how the revenue stream is generated and what barriers there may be to entry (Yock et al, 2015). In DF-HTA, the business model may not be fixed. Developers have the option to offer their technology (subject to local regulatory constraints) wherever the potential is greatest and to target patients and/or clinicians directly or to sell via national health services. For example, van Nimwegen et al (2017) used parents' willingness to pay for a diagnosis to calculate 'headroom' (valuing an estimated extension in life and/or improvement in quality of life at a given threshold value with an adjustment for the cost impact of the technology) rather than an explicit threshold for reimbursement as it was felt that the technology would be best suited to the private payer market. The business model adopted by the commercial developer or investor may differ across jurisdictions. Non-commercial developers may also need to consider commercial means of bringing their technology to market, as established biotech companies maybe best-placed to maximise the technology's potential.

8. Resources for analysis

In the early stages of development, in large companies, there may be a set of candidate technologies which could potentially be assessed using DF-HTA. As many of these potential technologies will fail (Annemans et al, 2000; Vallejo-Torres et al, 2008) resource-intensive approaches to HTA themselves may not

have a positive expected net present value. DF-HTA must compete for scarce resources, potentially displacing aspects of the research and development process. In addition, many medical devices, including tests, are developed by small and medium-sized enterprises and may be the sole product of that company (Craven et al, 2012). Such companies may have limited HTA experience and resources. This means that DF-HTA must deliver value within significant resource constraints. At the earliest stages of development, it is suggested that effort is focused on articulating and quantifying a value proposition (Vallejo-Torres et al, 2008; Rogowski et al, 2016). This could potentially be done using qualitative interaction with clinicians and users (Davey et al, 2011; Kluytmans et al, 2019) and simple quantitative methods such as headroom analysis (Chapman, 2013). This prima facie case can then be developed further as the development progresses when more resources may be available (Vallejo-Torres et al, 2008; Buisman et al, 2016b, Rogowski et al, 2016).

9. Stance of analysis

In this context, stance of analysis means the mindset adopted by the analyst in undertaking the assessment. The adoption of a positive rather than a normative economic stance of analysis is one of the fundamental features of DF-HTA, which has not previously been widely discussed. DF-HTA for commercial developers adopts a positive stance, as no value judgements are required (Culyer, 2010) and the analysis is focused on the maximisation of the developers' return on investment. For example, Hummel et al (2012) mentioned that the aim of their analysis was to 'support the future development' of the technology. Similarly, Kluytmans et al (2019) commented that much early HTA 'has a strong technology-focused or supply-driven character'. Developers start with the technology and part of the role of DF-HTA is to find a place where it can be successful. In this sense, DF-HTA has the character of a formative assessment i.e. an assessment to further the development. By way of contrast, use-focused HTA has the character of a summative assessment against a pre-determined set of criteria. Use-focused HTA adopts a normative stance; it involves judgements about what is good for society (Culyer, 2010).

10. Burden of proof

There are no guidelines about either methods to be adopted or the acceptable level of evidence required for DF-HTA, nor would such guidelines be appropriate. The process of DF-HTA is iterative; initial stages use whatever evidence is available and methods deemed appropriate by the analyst. The output from the HTA process informs the discussion between the developer and the analyst and takes any limitations in evidence and methods of assessment into account. For use-focused HTA, in many jurisdictions, there are clear guidelines as to what level and form of evidence the reimbursement agency or payer deems acceptable as well as how the assessment should be undertaken. For example, the National Institute for Health and Care Excellence, the reimbursement agency for England and Wales, prefers the evidence of health effects to come from randomised controlled trials directly comparing the intervention with one or more relevant comparators and has a comprehensive guide to methods (NICE, 2013).

4.2.2.2 How the list of features may be used

This first element of the framework, the list features of DF-HTA, could be used as an aide-memoire at the planning stage of a project, in initial discussions with developers. This would help to clarify essential features of the analysis in the mind of the analyst and ensure transparency between the developers and the analyst. Certain features of the framework may encourage discussions about features which would be unlikely to be discussed otherwise, such as the developers' underlying objective. It may also encourage a consideration of the wider decision space or alternative business models. Additionally, the list and questions could be used as a checklist for reporting to developers or in a published article to ensure that the characteristics of the analysis are transparent. Table 4-2 includes a summary of questions for consideration or discussion.

Table 4-2: Questions for consideration in DF-HTA

Feature of DF-HTA	Questions for consideration
Target audience	Who is the analysis designed to inform?
Underlying user objective	What are the developers ultimately trying to achieve through investment in development of a technology? On what basis will the developers decide whether and how it is worth continuing with the development of this technology?
Decisions HTA designed to inform	What decisions can the analysis inform?
Decision space	What are the possible uses of the technology? What are the most promising uses of the technology? Which of the potential use(s) should be targeted first?
Available evidence	What evidence is available? What is the best approach to estimating parameters in the absence of evidence?
Timing	What is the most appropriate form of analysis (if any) to do now?
Business model	What alternative business models are possible for this technology in target jurisdictions/indications?
Resources for analysis	What resources are available for analysis? What would be the most appropriate use of the resources?
Stance of analysis	How does the analyst ensure the study meets the needs of the developers?
Burden of proof	Are the methods and sources of parameter estimates appropriate for this level of resources and this stage of development? Has the analyst communicated any limitations of the approach with the developers?

DF-HTA – development-focused HTA, HTA – Health Technology Assessment

4.2.3 Conclusions

This chapter set out to provide a characterisation of HTA undertaken to inform developers. Ten features of DF-HTA are set out which can be used as an aide-memoire for analysts new to this work and as a checklist for reporting. Four of the features (target audience, decisions to inform, available evidence and timing) had been included in previous frameworks distinguishing early and mainstream HTA (Ijzerman and Steuten, 2011) or classical HTA (Pietzsch and Paté-Cornell, 2008). The remaining features (underlying user objective, decision space, business model, resources available for analysis, stance of analysis and burden of proof) were identified from the literature (see Appendices 3-5) with input and discussion with PhD supervisors.

Although previous authors have gone some way towards characterising DF-HTA (as part of "Early HTA") (Ijzerman and Steuten, 2011; Markiewicz et al, 2014; Pietzsch and Paté-Cornell, 2008), it was often conflated with other activities where evidence was scarce such as horizon-scanning (Ijzerman and Steuten, 2011) and the assessment of process or innovation from the perspective of the health service provider (Markiewicz et al, 2014; Dong and Buxton, 2006). Although these related activities may share some features with DF-HTA such as the timing and the lack of evidence, they differ significantly in important aspects of the work. In particular, the target audience for health service perspective work is healthcare decision-makers and the stance of analysis may be normative in nature. Authors associated with the MATCH collaboration in the UK set out a methodology for DF-HTA (Cosh et al, 2007; McAteer et al, 2007; Chapman, 2013) but did not attempt a comprehensive characterisation of this form of HTA. There was a recognition from this research group that this work, undertaken primarily for small and medium enterprises in the assessment of devices, was "a different animal" from use-focused HTA. For example, McAteer et al used the term "supply-side" HTA (McAteer et al, 2007). However, this is the first comprehensive attempt to set out the features of HTA to inform developers.

Formal validation of the list of features is hampered by the, understandably, limited number of published examples of DF-HTA, especially commercial examples. There is little incentive for developers to publish HTA studies and the need for commercial confidentiality creates a disincentive. This means that the body of published literature is skewed towards work funded by a public body and/or supported by translational research bodies. A recent useful article by Grutters et al (2019) highlighted this bias. It summarised 32 assessments of 30 non-drug technologies undertaken by their academic group in the Netherlands. Of the 32 studies, 30 were designed to inform developers and all but two were unpublished. All the developers were small or medium sized enterprises. The features described by Grutters et al (2019) supported the identified list of features characterising DF-HTA in the range of decisions to be informed and broad decision space. Timing of the assessments in this study ranged from idea screening, through concept development, pre-market and market access. 50% of

the technologies assessed were already available on the market so the timing is potentially a little later than was envisaged in the characterisation of DF-HTA.

For analysts outside of large device or drug companies, new to this work, this first element of the framework of DF-HTA provides a clear introduction to the features of DF-HTA and will guide their discussions with developers to ensure both parties are clear on the distinct nature of this work. It should also improve the transparency of any published DF-HTA if the features of each study are reported. For the wider academic community, this initial characterization of HTA to inform developers may provoke debate among practitioners about the nature of this work and the accuracy of the characterisation of DF-HTA. Further research which would be of use include studies examining the features of DF-HTA in the commercial context and empirical studies applying this element of the framework. The different features of DF-HTA also necessarily impact on methods adopted. This has been explored for Early HTA in the academic literature (Markiewicz et al, 2014) but the boundaries of DF-HTA were not clearly established.

4.3 Contribution of this chapter

The main contribution of this chapter is to develop a characterisation of DF-HTA in ten features. Only four of these features had previously been identified explicitly. The characterisation may help analysts new to working with technology developers to understand the different requirements of development-focused projects compared to use-focused work, with which they are generally more familiar.

The features of DF-HTA set out in the framework drive important differences in the methods of analysis used. The process and analytical methods used in DF-HTA are the subject of the following chapter.

5 Developing a framework of HTA to inform developers: part two - process and methods

5.1 Introduction

Chapter 5 addresses research questions 2 and 3:

2. What is the process of HTA undertaken to inform developers?

3. What are appropriate methods of HTA to inform developers of health technologies and how do these methods link with the particular features of HTA to inform developers?

The previous chapter proposed that HTA undertaken to inform developers (DF-HTA) has ten distinct features. This chapter builds on that contention and looks at the process and methods of DF-HTA. The chapter has three elements:

- the development of a generic process of DF-HTA (Section 5.3.1). This focuses on high level activities of DF-HTA and the information flows connected with those activities.

- the identification and categorisation of analytic methods of DF-HTA used to generate the information flowing between activities (Section 5.3.2).

- a description of analytic methods together with examples and critique of the methods (Section 5.3.3).

5.2 Methods

The full body of papers identified in the literature review described in Chapter 3 was reviewed to identify frameworks for HTA to support developers of medical technologies (Section 5.3.1). Inclusion criteria were that the authors of the study explicitly set out a framework and that it included more than one activity of DF-HTA. Aspects of the identified frameworks were reviewed and discussed with thesis supervisors and a generic process of DF-HTA was suggested and justified. The full body of development-focused papers identified in the literature review informed the second and third elements of this chapter:

identifying and categorising methods of DF-HTA (section 5.3.2). The methods described or applied in included papers were extracted along with the aspect of the information flow they were supporting. Terminology was streamlined where similar terms were describing essentially the same method. For the third element, the link between the features of DF-HTA and the method of DF-HTA, authors of the papers generally made no reference to the features of DF-HTA so the connection between the feature and the approach taken had to be deduced. The one feature which was referred to by some authors was resource constraint.

5.3 Results

5.3.1 Developing a generic process of DF-HTA

5.3.1.1 Frameworks of DF-HTA from the literature

From the body of papers identified in the literature review described in Chapter 3, eight papers presented frameworks of HTA to support developers of medical technologies (Cosh et al, 2007; Retel et al, 2008; Davey et al, 2011; Ijzerman and Steuten, 2011; Rogowski et al, 2016; Markiewicz et al, 2017; Borsci et al, 2017; de Graaf et al, 2018). These are set out in Table 5-1. All frameworks included clinical and economic value assessment. The frameworks varied in the extent to which they included other components, such as market/business case, the explicit involvement of users/other stakeholders and post-market considerations. A key consideration is the extent to which these other components form part of DF-HTA or whether DF-HTA is focused on the properties and evidence related to the technology itself. The different components included in each framework are discussed in the remainder of this section. In section 5.3.1.2 a proposed generic process of DF-HTA is set out and justified.

Table 5-1: Identified frameworks of DF-HTA

First author	**Cosh**	**Retel**	**Davey**	**Ijzerman**	**Rogowski**	**Borsci**	**Markiewicz**	**De Graaf**
Date	**2007**	**2008**	**2011**	**2011**	**2016**	**2017**	**2017a**	**2018**
Technology type	**Generic**	**Nanotechnology in oncology**	**Devices**	**Generic**	**Generic**	**Devices**	**Devices**	**Biomarkers**
Component of decision framework	Strategic considerations		Value proposition	Business case analysis	Value proposition		Market assessment	
	Clinical problem definition	Clinical – safety, efficacy, effectiveness		Clinical case analysis		Problem structuring – stakeholders, operational, challenges, gold standard	Clinical context assessment	Clinical impact
		Patient-related – ethics, acceptability, psychosocial reactions, patient centeredness, patient-related juridical aspects					Stakeholders analysis	
		Organizational – adoption, implementation, accessibility/equity, skills/routines/logistics, juridical, education/training				Evaluation components (workflow, user experience, unmet needs)		
						Space of adjustments – scenarios of use		
	Headroom	Economic - Cost-effectiveness	Headroom	Headroom		Device definition - initial modelling of scenarios	Financial and health economics evaluation	Headroom
	Return on investment		Business model including competitive strategy	Return on investment				
	Further economic evaluation			Early health economic evaluation	Establishment of evidence	Modelling and clinical trial		
		Scenario/Roadmap – Diffusion scenario			Challenges of information and motivation			

One framework included just clinical and economic aspects (de Graaf et al, 2018). The aim of de Graaf et al's study was to demonstrate an approach to the very early assessment of potential biomarker technologies without extensive modelling being required so economic analysis was restricted to the estimation of commercial headroom using a simple formulaic approach.

Four of the frameworks included an aspect of strategic analysis (Cosh et al, 2007; Davey et al, 2011; Ijzerman and Steuten, 2011; Markiewicz et al, 2017a). For example, the first stage of the framework put forward by Cosh et al (2007) included strategic considerations (such as the fit of the new technology with the developer's competencies and assessment of the competitive environment) prior to the definition of the clinical problem. Ijzerman and Steuten (2011) and Markiewicz et al (2017a) also included strategic activities in their frameworks but described them as 'business case analysis' and 'market assessment' respectively. Davey et al's (2011) framework saw clinical problem definition and early health economic analysis feeding into the development of a six-part business model (taken from Chesbrough, 2006) comprising value proposition, market segment, value chain required to create and distribute the technology, revenue generating mechanisms and profit potential, position of firm within value network including complementors and competitors and formulating competitive strategy.

Two frameworks (Retel et al, 2008; Borsci et al, 2017) had a particularly strong focus on the inclusion of user and other stakeholder perspectives. Retel et al (2008) set out a framework for the evaluation of innovations in nanotechnology in oncology based on the Constructive Technology Assessment (CTA) approach (Douma et al, 2007), which has primarily been utilised in the Netherlands for the assessment of non-pharmaceutical technologies (Bartelmes et al, 2009). This approach sees health economic considerations embedded in a comprehensive research and development methodology. There is a strong emphasis on patient-related factors such as ethics, acceptability and psychosocial reactions and organisational factors such as what changes in skills, routines and logistics would be required for the technology to be successful (Retel et al; 2008). For example, a constructive technology assessment conducted alongside a clinical trial of a prognostic signature in breast cancer investigated patient acceptability of the

test. It found that patient satisfaction with the test was lower when the results of the test (i.e. low or high risk of progression) were given on a separate occasion to risk of progression according to previous standard of care based on clinical and pathological factors. This was particularly the case when results were discordant. The same assessment included analysis of the impact on diffusion of physician resistance to the test due to uncertainty in the evidence base. This scenario was considered unlikely at the outset of the assessment but proved to be realistic as the validity of a test based only on retrospective validation was called into question by Dutch physicians. This approach would appear to be particularly useful in co-development situations where the developer and a healthcare organisation are working collaboratively to evaluate and potentially modify a technology in the later stages of development. Borsci et al (2017) presented a framework which combined health economics and human factors approaches (Carayon et al, 2014). Human Factors or ergonomics is defined by the International Ergonomics Association (undated) as:

> 'the scientific discipline concerned with the understanding of interactions among humans and other elements of a system'.

Borsci et al's framework was developed at one of the UK National Institute for Health Research (NIHR), Diagnostic Evidence Co-operative (DEC) centres, now replaced by Medtech and In vitro diagnostics Co-operatives (MICs) (NIHR, 2019). The framework suggested that human factors be incorporated into health economic modelling. For example, workflow analysis using task analysis, observation and interview could feed into both design of the technology and provide contextual information for the health economic analysis (Borsci et al, 2017).

Retel et al (2008) and Rogowski et al (2016) both included components which considered post-market aspects of a technology's life cycle. Retel et al (2008) included scenario analysis or road-mapping of different diffusion scenarios as a final component and Rogowski et al (2016) included challenges of information and motivation as the final component in their three-stage framework. These challenges primarily refer to the misalignment of incentives in technology innovation and diffusion. The initial and continued success of a technology depends upon the extent of its diffusion.

5.3.1.2 A proposed generic process of DF-HTA

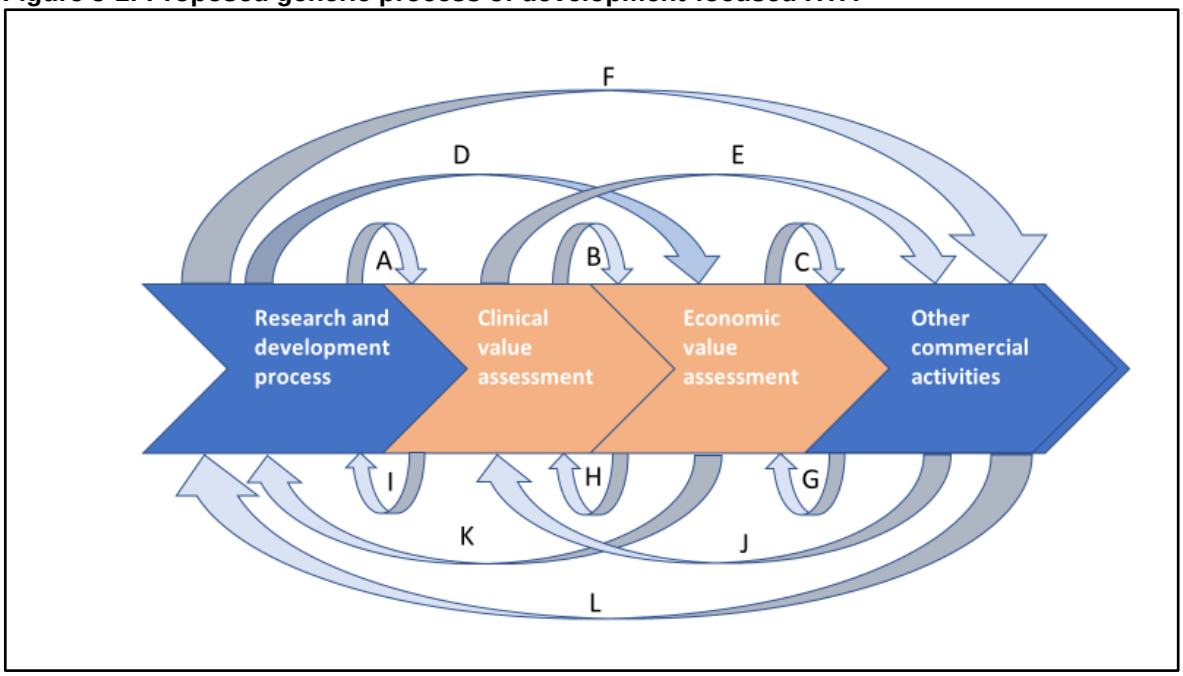
The proposed generic process (Figure 5-1) adopts as its core the two key aspects of DF-HTA identified in all the frameworks found in the literature – clinical value assessment and economic value assessment. DF-HTA is restricted to activities which involve the **assessment** of the technology. These core activities are shown in orange in Figure 5-1. The clinical value assessment considers what impact the technology might have on clinical practice and ultimately upon health (and wider social) outcomes. The economic value assessment builds on the clinical value assessment to consider the economic impact of changes in healthcare resource use and other economic value drivers such as productivity effects. The clinical value assessment must always precede the economic value assessment as the latter is dependent upon the former. The arrows connecting the clinical and economic value assessment indicate the iterative nature of the assessments with both aspects revisited numerous times during the course of the development. For the purposes of this thesis these core aspects comprise DF-HTA.

The remaining components of Figure 5-1 are the research and development process and other commercial activities, both shown in blue. These two components are not considered to form part of DF-HTA although they are closely associated with it. The research and development process includes the generation of the concept for and the design of the technology, as well as the generation of evidence of its safety and clinical effectiveness. Other commercial activities includes the consideration of potential markets, pricing, volume of sales, fixed and variable costs of production, estimates of net margin and ultimately whether the continuing development is likely to deliver a positive financial or societal return on investment. It may also involve consideration of portfolios of candidate technologies.

Along with the two core aspects of DF-HTA shown in orange in Figure 5-1, some of the frameworks discussed in 5.3.1.1 included consideration of strategic/market issues, stakeholder/user interaction and post-market issues. Consideration of all these issues forms part of the proposed generic process although not necessarily as development-focused HTA. Strategic/market issues would form part of other commercial activities. Stakeholder/user interaction

may form part of any of the components depending on what the interaction is seeking to achieve. It would be part of the research and development process where it seeks to inform technology design. It would be part of the clinical or economic value assessments if input from clinicians and patients was sought, for example, in order to map clinical pathways. It would be part of other commercial activities if stakeholders were consulted in an attempt to estimate the size of the market. Post-market issues, particularly around encouraging adoption of the technology, would form part of other commercial activities but may be informed by the clinical and economic value assessments.

Figure 5-1: Proposed generic process of development-focused HTA



Core activities of development-focused HTA are shown in orange, related activities are shown in blue. Arrows indicate the information flows between the different activities.
A – Evidence of clinical effectiveness.
B – Evidence of clinical effectiveness for technology and comparator, contextual information to allow pathway modelling and diffusion information to inform modelling and/or scenario analysis.
C – Commercial headroom, pricing thresholds, evidence gaps, insight on potential impact of diffusion related parameters.
D – Cost profile of technology.
E – Potential market size.
F – Potential indications to target.
G - Target markets and indications for scenario selection.
H – Markets/indications with most economic potential.
I – Evidence gaps, threshold technology performance to improve clinical effectiveness, insights into contextual aspects requiring consideration in design.
J – Target markets and indications.
K – Threshold technology performance for cost effectiveness, evidence gaps.
L – Commercial headroom, threshold for costs of production and development.

The presentation of Figure 5-1 suggests a time element to the framework. This is true only to the extent that the clinical value assessment must necessarily precede an economic value assessment as the change in clinical pathways forms the basis of any decision model. The research and development process has been put first as in practice, the technology (i.e. research and development) often precedes any DF-HTA (Kluytmans et al, 2019). In some situations, commercial considerations or an identified clinical need will drive the research and development process. The arrows and caption to Figure 5-1 demonstrate the complexity of the links between each of the components and the iterative nature of the process.

5.3.2 Categorising methods by development-focused HTA activity

5.3.2.1 Classification of methods from the literature

The starting point for the classification of methods was the literature identified in the review. As described in Section 3.2.4, methods identified in the search included many which contribute to the research and development process or to other commercial activities rather than to either the clinical or economic value assessments. For the sake of completeness, all methods were classified into one of the four components of the generic process (see Appendix 6). Detailed discussion is reserved for the methods which contribute to the components of DF-HTA (i.e. clinical value assessment and economic value assessment) including methods of providing estimates to feed into either the clinical or the economic value assessment. It should be noted that Table 5-2 does not aim to set out a comprehensive list of methods for any of the components as this was not the purpose of the literature review. It only includes methods identified previously as methods of early HTA including HTA to inform developers.

The literature review identified methodological studies describing methods of development-focused HTA (n=43), studies applying methods of DF-HTA (n=25) and studies which involved a combination of the above factors (n=20). Six studies listed and described a selected group of methods (Miller, 2005; Hartz and John, 2008; Bartlemes et al,2009; Ijzerman and Steuten, 2011; Mikudina and Redekop, 2014; Markiewicz et al, 2014). A seventh source, (Graziadio et al, 2020) was added at a late stage of thesis drafting as the methods described were relevant. The categorisation process is shown in detail in Appendix 6 and the resulting parsimonious list of analytical methods which may be useful for DF-HTA is set out in Table 5-2. Definitions of the methods are given in Table 5-3. The paragraphs following the tables explain the methods and provide examples of their application from the retrieved studies.

5.3.2.2 Overview of methods of DF HTA

Table 5-2: Analytic methods of DF-HTA

Analytic methods (Sub-types)	What the method involves	Activity of DF-HTA	Information provided	Adaptations to the method for the purposes of DF-HTA
Care pathway analysis	Mapping of existing and potential clinical pathways	Clinical value assessment Economic value assessment	Visual representation of existing and potential care pathway Market size Technology positioning Potential market gaps Identification of comparators	Literature review less rigorous (reliance on small range of authoritative sources such as guidelines/HTA reports) Less formal methods of expert elicitation
Qualitative methods of stakeholder consultation (e.g. interviews, focus groups and surveys)	Consulting clinical experts, patients or other stakeholders	Clinical value assessment Economic value assessment	Articulation of value propositions Evidence of clinical effectiveness for technology and comparator Insight into contextual aspects requiring consideration in design. Barriers and facilitators for diffusion to inform economic model structure and/or scenarios for analysis.	Less formal methods used Smaller numbers consulted Limited number of settings
Literature review	Review of publicly available information and the academic literature	Clinical value assessment Economic value assessment	Existing care pathways Evidence of clinical effectiveness for technology and comparator Evidence of costs and utilities for comparator and potentially for technology Identification of current and potential comparators Barriers and facilitators for diffusion to inform economic model structure and/or scenarios for analysis.	Less rigorous methods than systematic review. Reliance on small range of authoritative sources such as guidelines/HTA reports at earliest stages.
Multi-criteria decision analysis	Comparing alternatives using weighted criteria	Clinical value assessment Economic value assessment	Evidence of clinical effectiveness for technology and comparator Pricing Preferences of decision makers/users for attributes of a new technology	Use of simple, transparent sub-type of multi-criteria decision analysis unless well resourced.
Discrete choice experiments	Pairwise comparisons to reveal preferences for attributes	Clinical value assessment	Evidence of clinical effectiveness for technology Preferences of decision makers/users for attributes of a new technology	Too complex for use in DF-HTA unless well resourced.

Analytic methods (Sub-types)	What the method involves	Activity of DF-HTA	Information provided	Adaptations to the method for the purposes of DF-HTA
		Economic value assessment		
Expert opinion (e.g. Delphi Panel)	Consulting experts using qualitative methods. Delphi panels involve two rounds of consultation.	Clinical value assessment Economic value assessment	Identification of comparators Existing and potential care pathways Barriers and facilitators for diffusion to inform economic model structure and/or scenarios for analysis.	Less formal methods used Smaller numbers consulted Limited number of settings
Expert elicitation	Consulting experts using quantitative methods to derive point estimates and/or probabilities with range	Clinical value assessment Economic value assessment	Estimates of clinical effectiveness for technology and comparator Estimates of costs and utilities for comparator and potentially for technology	Less formal methods used Smaller numbers consulted Limited number of settings
Cost-effectiveness analysis (cost utility analysis cost consequence analysis cost benefit analysis cost minimisation analysis estimation of headroom, budget impact analysis)	Comparing the costs and health outcomes of existing and potential care pathways	Economic value assessment	Commercial headroom Target thresholds for development costs, pricing and clinical effectiveness/test performance Evidence gaps Reimbursement potential Potential cost savings Budget impact	Simple models Multiple scenarios Shorter time horizons Intermediate outcomes Use of cost minimisation and cost consequence analysis One-way sensitivity analysis
Value of information analysis (Expected value of perfect information, expected value of partial parameter information, expected value of sampling information)	Estimating the value of further research to society or a commercial developer using a probabilistic decision analytic model	Economic value assessment	Estimates of the societal or commercial value of further research Identification of influential parameters for reducing uncertainty in cost-effectiveness Estimate of optimum sample sizes	Too complex for use in DF-HTA unless well resourced.

DF-HTA – development-focused HTA. See Appendix 6 for source articles of methods. Definitions of methods of development-focused HTA are given in Table 5-3.

Table 5-3: Definitions of analytic methods of development-focused HTA

Method	Definition	Source of definition
Care pathway analysis	"Care pathway analysis is the method of modelling a care pathway in a healthcare system. It is a type of systems or process analysis. The resulting model is shown graphically as a systems diagram or map of the services provided to a typical patient"	Graziadio et al, 2020
Qualitative methods of stakeholder consultation (e.g. interviews, focus groups and surveys)	Focus groups – "Method based on asking questions in an interactive group setting where participants are free to talk with other group members, in order to gather the information about perceptions, opinions, beliefs, and attitudes towards a topic/product of interest."	Markiewicz et al 2014
	Workshops – "a method based on creating a setting where participants can benefit from focused interaction with each other. The goal is to discuss and to exchange experiences on a range of relevant topics in medical devices development, during facilitated sessions."	Markiewicz et al 2014
	Interviews – "The qualitative research interview is performed to describe and understand the meanings of central themes in the medical devices development."	Markiewicz et al 2014
	Survey research – "The broad area of survey research encompasses any measurement procedures that involve asking questions of respondents (e.g. a short paper and pencil feedback form to an intensive one-on-one in-depth interview) to get the user involvement in the development process."	Markiewicz et al 2014
Literature review	"Reviewing a body of text to come up with the critical points of current knowledge including substantive findings as well as theoretical and methodological contributions to a particular topic of interest in the medical device's development." "Topic, product or field of interest may refer to different parts of analysis and actions performed during the development of the medical devices, in the fields like: applications, patient populations, patients preferences, usability, cost-effectiveness, etc"	Markiewicz et al 2014
Multi-criteria decision analysis	"A series of techniques aimed at supporting decision makers faced with making numerous and sometimes conflicting evaluations. Multi-criteria decision analysis aims at highlighting these conflicts and deriving a way to come to a compromise in a transparent process."	Ijzerman and Steuten, 2011
Discrete choice experiments	Choice-based conjoint analysis (Discrete Choice Modelling) – "Method based on the choice experiments: a test person is confronted with a small number of options sampled from a parameterized space, and has to choose his preferred option."	Markiewicz et al 2014
Expert opinion	Expert opinion – "studies that aim to draw forth the opinions or beliefs of experts expressed in a qualitative format"	Iglesias et al 2016
Expert elicitation	Expert panels/elicitation (e.g. Delphi method) – "A highly structured technique in which selected experts provide their assessment of likely future outcomes of implementing new medical device by responding to several rounds of questions."	Markiewicz et al 2014
Cost-effectiveness analysis (cost utility analysis cost consequence analysis cost benefit analysis cost minimisation analysis estimation of headroom, budget impact analysis)	Cost-effectiveness analysis "compares the relative costs and outcomes (effects) of two or more courses of action to find the best alternative activity, process or intervention that minimises resource use to achieve the desired result".	Markiewicz et al 2014
	Cost-utility analysis – "a method aimed to estimate the ratio between the cost of a health-related intervention and the benefit it produces in terms of the number of years lived in full health by the beneficiaries."	Markiewicz et al 2014
	Cost-consequence analysis – "is a form of economic evaluation where disaggregated costs and a range of outcomes are presented to allow readers to form their own opinion on relevance and relative importance to their decision-making context"	Drummond et al 2015
	Cost-benefit analysis – "a systematic process for calculating if a medical device development project is a sound investment/decision, and to provide a basis for comparing projects."	Markiewicz et al 2014

Method	Definition	Source of definition
	Cost-minimisation analysis – "compares interventions based solely on their net cost"	Hunink et al 2014
	Headroom "is a relatively simple threshold approach … that estimates the maximum amount that a technology could cost and yet still be considered cost-effective."	Redekop and Mikudina, 2013
	A budget impact analysis provides a measure of the affordability of adopting a medical device by quantifying the effects on the budget of a healthcare provider.	Graziadio et al, 2020
Value of information analysis (Expected value of perfect information, expected value of partial parameter information, expected value of sampling information)	Value of Information analysis (VOI) – "an analysis aimed at presenting the amount a decision maker would be willing to pay for information prior to making a decision during the medical device development to avoid uncertainty."	Markiewicz et al 2014
	Expected Value of Perfect Information (EVPI)... "reflects the discrepancy between the current information position and a position of no perfect information (no uncertainty)" "population EVPI can provide a measure of the maximum return of future research, placing an upper limit on the societal costs of it".	Steuten et al 2013
	Expected Value of Partial Parameter Information (EVPPI) – "informs us for which specific consequences of the technology (e.g. impact on utilities, costs or health status) more information is needed to make a less uncertain decision in the future, again offset by the costs of collecting that further information."	Steuten et al 2013
	Expected Value of Sample Information (EVSI) – "estimates the expected VOI that could be gathered from a sample of a given size *n* within a particular study design, over a specified time period".	Steuten et al 2013

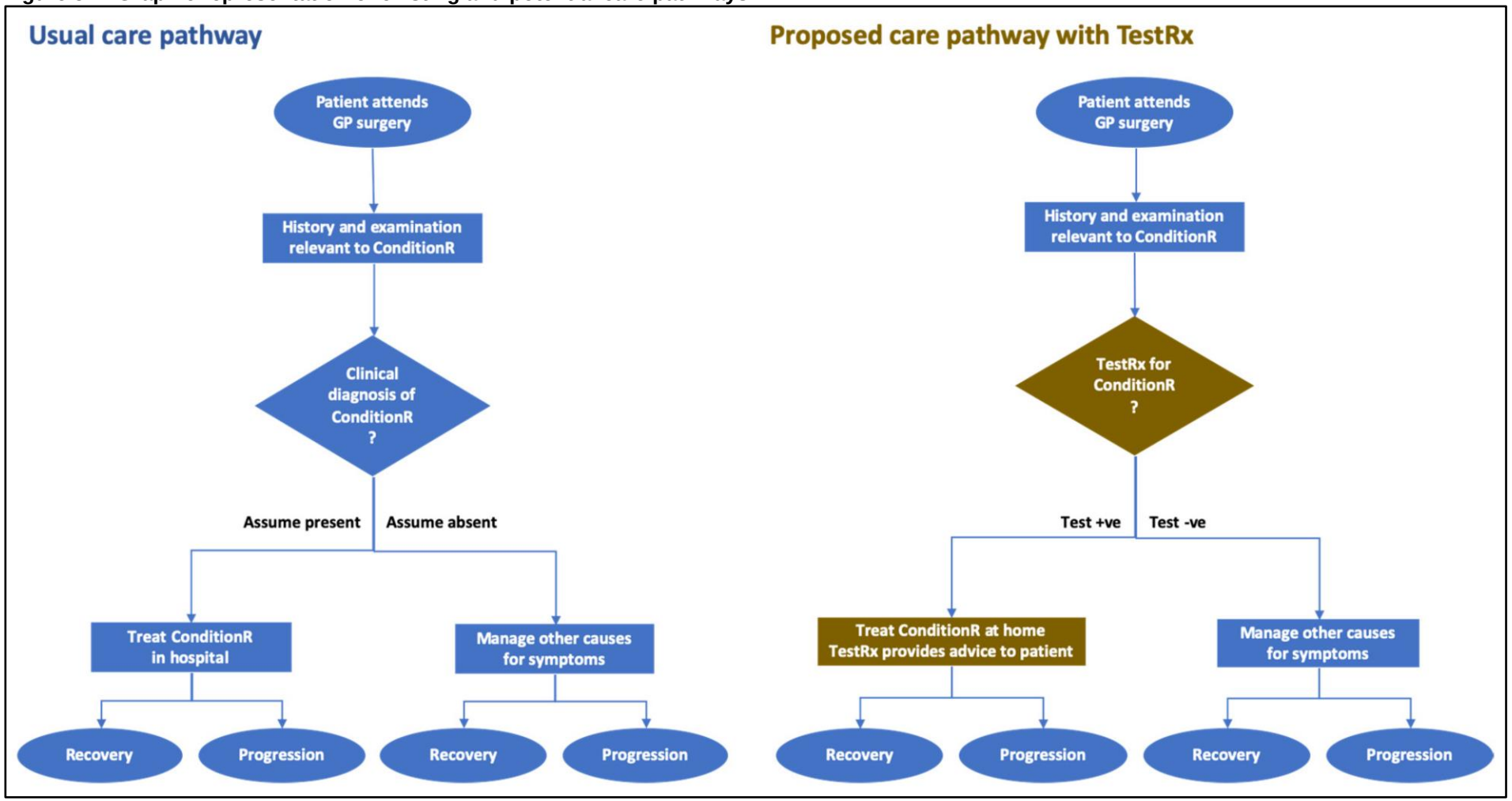
5.3.3 Methods of DF-HTA

5.3.3.1 Care pathway analysis

Often development of health technologies is 'technology driven' (Martin et al, 2008; Weiser, 2013). An example of technology-driven development would be where engineers at a university develop a new imaging modality without having a single clinical application in mind. The starting point for any DF-HTA then involves positioning the technology in the relevant care pathway in order to determine whether it is likely to add value. Care pathway analysis involves the representation of clinical care pathways, often in a visual format, although a narrative representation may supplement the process diagrams (Graziadio et al, 2020). Information to populate the care pathway analysis can be derived from literature review and qualitative methods of stakeholder consultation such as interviews, surveys or focus groups. Care pathway analysis is the necessary first step of an economic evaluation as the existing and potential care pathways form the alternative strategies to be modelled. Multiple care pathways may need to be modelled as existing care pathways may vary within and across jurisdictions and the technology may be able to be used at multiple positions in the pathway.

Figure 5-2 illustrates a visual representation from a generic care pathway analysis for a new test technology which Graziadio et al (2020) have named Test Rx. Ovals represent inputs and outputs, lines with arrows represent paths and directions, rectangles represent processes and diamonds decisions. The brown shapes in the proposed care pathway indicate where the new technology changes the clinical pathway. The new test replaces clinical decision making and allows patients testing positive to be managed at home rather than admitted to hospital. Although the ultimate outcomes remain the same (recovery or progression) the new technology may change the proportions of patients with these outcomes. These changes form the basis of the value proposition of the technology and inform both modelling and evidence generation strategies.

Figure 5-2: Graphic representation of existing and potential care pathways.



Reproduced from Graziadio et al, 2020 (no permission required)

Although the authors do not offer a step by step guide to care pathway analysis they provide the eleven recommendations set out in Table 5-4.

Table 5-4: Recommendations for care pathway analysis

1	Work with the manufacturer to articulate opportunities for improvement in the clinical pathway using the technology
2	Work collaboratively with a team of stakeholders including manufacturers, healthcare payers, providers and patients
3	Review national and local guidelines
4	Obtain frequent feedback from clinicians and patients (consider ethics in patient interaction, use surveys, interviews and focus groups)
5	Summarise current clinical evidence (consider analytic validity, clinical validity and clinical utility)
6	Articulate the current clinical scenario (problem, setting, recommended management)
7	Describe the technology, indications for use, strengths and limitations, who will use it and where?
8	Describe how the information produced by the product will guide management decisions
9	Obtain comments on usability and potential utility from stakeholders
10	Describe outcomes (clinical and economic) that are expected
11	Visit the places (preferably with an example of the technology) where the device and the information produced will be used and talk to potential users of the technology or the information it produces.

Adapted from Graziadio et al, 2020

Examples of care pathway analysis

No examples of care pathway analysis were identified in the literature review. Graziadio et al (2020) refer to one example of a similar methodology 'patient-pathway analysis' (Hanson et al, 2017). Rather than assessing how a new technology may impact existing care pathways, patient pathway analysis aims to to understand the alignment between how patients access clinical services and how services are provided. The context in Hanson et al (2017) was planning of tuberculosis services using survey data. The outcome of the analysis illustrates gaps in services and may indicate useful starting points for technology innovation.

Strengths of care pathway analysis

- Undertaking care pathway analysis involves a consideration of the placement of the technology in an existing clinical pathway which requires developers to explicitly consider the clinical utility of their technology compared to existing and emerging comparator technologies.

- Early engagement with potential payers, clinicians and patients may provide useful input to the design process or reconsideration of the target market.

- Visual representation of pathways allows shared understanding between different groups of stakeholders, facilitating discussion and consensus-building.

- Care pathway analysis facilitates the articulation of value propositions which can feed into marketing strategies, business plans and evidence generation strategies.

- Multiple pathways can be represented and the analysis can be updated on an iterative basis as the development progresses.

- The existing and potential clinical pathways identified form the basis of the health economic analysis.

Limitations of care pathway analysis

- There is a lack of uniform terminology or methodology for care pathway analysis (Graziadio et al, 2020). Although there are few explicit examples of care pathway analysis, the approaches described under qualitative stakeholder consultation have many common features with the care pathway analysis.

- Diagrammatic representations of care pathways may not capture all aspects of a technology's value such as more rapid diagnosis or improved quality of life for a patient. Providing these elements are captured in narrative form they can be incorporated into the health economic analysis and used in marketing and business plans.

- The proportions of patients taking different paths are not captured by care pathway analysis nor are the proportions of patients experiencing the alternative outcomes. These can, however, be built into the health economic analysis.

5.3.3.2 Qualitative methods of stakeholder consultation

Several applied studies identified in the review have described clinical value assessment using qualitative methods of stakeholder interaction although there is no consistency in the terminology used. For example, Davey et al (2011) referred to 'field research' to determine 'the value proposition', Markiewicz (2017b) referred to a 'stakeholder consultation exercise' and 'iterative stakeholder involvement', Kluytmans et al (2019) referred to 'early assessment of a proof of problem', Wieser (2013) referred to a 'health needs assessment', Wissing (2012), a 'clinical case scenario analysis' and Graziadio et al (2020) to 'care pathway analysis'. Despite the difference in terminology, these methods seem to share a common approach, although this is not always explicit and not all steps are undertaken in all studies. The approach involves up to five steps. The five-step approach is set out in Table 5-5 and two examples follow the table to exemplify the approach. Following the examples, the strengths and limitations of this approach are discussed.

Table 5-5: Five-step approach to CVA using qualitative stakeholder interaction

Step	Description
1	Identify the most favourable potential application/s for the technology using information from the literature, developers and in some studies stakeholder input
2	Identify the claims about the new technology. What are its advantages over existing and emerging technologies? Informants are generally the developers. Discussions could be structured using strategic tools such as PEST or SWOT (Chapman, 2013).
3	Identify the stakeholders. Generally, the potential users of the new technology and/or decision-makers involved in the purchasing decision. Often limited in number and/or range of settings.
4	Elicit stakeholders' views about current clinical pathways and the claims made for the new technology. Various methods may be used to elicit views including interviews, focus groups and surveys.
5	Develop implementation scenarios. Note barriers and facilitators for implementation.

CVA – clinical value assessment, PEST – political, economic, social, technological (Sammut-Bonnici and Galea, 2015), SWOT – strengths, weaknesses, opportunities, threats (Madsen et al, 2016)

Examples of qualitative stakeholder interaction

Davey et al (2011) is an example of a study covering steps two to four of the five step approach. The study aimed to set out the value proposition of a flexible stent in order to encourage funders to invest in clinical trials. The authors identified target clinicians likely to use the new technology who were leading consultants from the vascular surgery and interventional radiology departments

in two different centres in the UK. They held meetings in which the consultants were asked about clinical treatment routes, disease context, current treatments and challenges with current stent designs. In this study, the developers' claims for the technology were clarified by the stakeholder consultation allowing a three-part value proposition to be developed. No information was provided in the study about the number of stakeholders involved or about the precise methods of developing the value proposition. It was reported as having been a useful exercise for the developers as 'it confirmed the potential for an innovative device in the treatment of peripheral arterial disease and the requirement for investment in this area'. The authors also felt that the exercise allowed the developers to 'enhance marketing of the attributes of the technology that are likely to provide the greatest clinical benefit' and 'improve linkages in the value chain and value network by strengthening the relationships an organisation has with the users of the technology.'

Markiewicz (2017b) is a further example of a purely qualitative method covering all the steps in Table 5-5. This was a complex user consultation exercise over two stages of development, using a range of methods to consult a range of informants about an evolving point of care (POC) test of kidney function. In the first stage of the study, at 'targeting product' stage, a brainstorming session with the developers elicited the 'claims' for the technology and potential uses. The test was designed to detect sodium in urine and potassium, phosphate and calcium in blood and was aimed at patients at risk of kidney failure and their clinicians. The test was portable and had a quick turnaround time so would be suitable for daily screening. The developers claimed the test could result in reduced hospitalisation from more frequent testing and that the test was more accurate than current standard. Three potential implementation areas were identified 1) at home as a self-monitoring tool, 2) as a POC test in a GP's office, or 3) as a POC test in the hospital. A literature search had identified four populations at increased risk of developing kidney disease and the developers selected patients with diabetes mellitus type 2 as the most promising group to target due to the prevalence of kidney disease in that population and the size of the population.

40 people (22 patients, seven GPs, five radiologists and six radiographers) were recruited using snowball sampling across the Netherlands and took part in semi-structured interviews and completed questionnaires to elicit current practice in screening and monitoring and satisfaction with current care. The proposition of the new test was outlined and the needs and wants with such a test were elicited along with perceived clinical and ease of use benefits for both the specialist and the patient. Patients were asked about preferences for frequency, way of contact with GP, responsibility for acting on results, cost to the patient and health benefits, influence of the test on relationship with doctor and whether they felt capable and motivated to manage their disease. Patients indicated that they would prefer testing in the GP's office but GPs rejected frequent testing in this population as kidney disease develops slowly and current testing was adequate. GPs also indicated that any POC test would need to measure creatinine as this was the most important pathological sign of kidney failure. Radiographers and radiologists identified two groups who may benefit from rapid testing in hospital and six implementation scenarios were developed in different locations, populations and indications. These scenarios were explored in a larger sample of radiologists and radiographers and one was considered useful. This single scenario was rejected by the developers as the market was not sufficiently large and it would not be feasible to price the new device to compete with hospital laboratories.

In the second stage, the 'early proof of concept stage', the focus of the development had changed at this point to target patients with existing kidney disease who were either pre-dialysis or in dialysis and the detection of creatinine had been added to the functionality of the test. The developers felt the most potential for the test was in home monitoring. To explore the potential for home monitoring in this population, semi-structured interviews were held with seven key informants – one GP, one nephrologist, one pre-dialysis nurse, two patient representatives, one insurance company representative and one scientific researcher with experience of tele-health. Two scenarios discussed: patients monitoring at home and GPs monitoring in their offices. Following the key informant interviews, four scenarios were developed with manufacturers in a brainstorming session. Scenarios differed with the levels of self-monitoring responsibility, amount of contact with doctors and

reimbursement schemes. These scenarios were evaluated by 17 patients through a web-based questionnaire. The final preferred scenario was self-monitoring by the patient who then contacts the specialist if results deviate from expected levels.

Strengths of qualitative stakeholder interaction

- Where development is technology driven, developers may not have adequately considered whether a clinical need exists or precisely where the need is and how the technology must perform to meet the need. These consultation exercises explicitly explore where the need is and whether the technology is likely to fulfil it.

- Stakeholder interaction helps developers to be explicit about their value proposition which may help to raise finance and enhance marketing efforts (Davey et al, 2011).

- Negative findings can be used to redirect design effort at a relatively early stage of development to avoid costly late changes. For example, developers of a POC test in kidney disease were able to include detection of creatinine in order to improve the probability that the test could be used for screening (Markiewicz et al, 2017b). Developers of a surgical instrument were able to redirect attention to another application when stakeholder interaction revealed that there was little room for improvement in meniscus surgery (Kluytmans et al, 2019).

- Scenario development is useful for subsequent modelling exercises as it focuses modelling on areas where there is a realistic need. Stakeholder consultation may also provide insight into other aspects to be taken account of in clinical and/or cost effectiveness modelling such as the inclusion of patient and clinician preferences and behaviours.

- Stakeholder interaction may be a useful precursor to economic methods as unless a technology is likely to be cheaper than the current standard of care there may be no point continuing to economic assessment unless clinical value can be demonstrated.

Limitations of qualitative stakeholder interaction

- In the applied examples identified in the literature review, the number of stakeholders involved was limited to a small number and settings were generally single country. It is possible that stakeholders in other settings would have differing views, particularly if healthcare systems are differently organised or resourced. It is important, therefore, to see limited exercises in stakeholder interaction as informative but not conclusive. Negative findings may require an extension of the settings or indications explored rather than a decision to stop developing the technology. Positive findings should also be seen as informative rather than conclusive as more detailed evidence of clinical benefit is likely to be needed to persuade some decision-makers to adopt the technology.

- Many small developers would not have the capacity to undertake extensive user consultation exercises and developers may have difficulty gaining access to clinicians and other user groups (Weiser, 2013).

- It is difficult to distinguish between stakeholder consultation to inform design and consultation to inform assessment of the technology (see section 3.2.4). Indeed, Borsci et al (2017) presented a framework (discussed in section 5.3.1.1) which links health economic assessment and human factors/ergonomics as the assessment considers whether there is a clinical need and human factors/ergonomics considers what the characteristics of the technology needs to be in order to fulfil the need.

- Markiewicz et al (2017b) found that negative evidence from stakeholder interaction was not sufficient to influence the developers' decision about how to target the technology. In addition, the potentially useful scenario which was identified was rejected by the developers as there was insufficient potential for sales volume at feasible prices. These factors suggest that the analysis may not be considered useful by the developers. It should be remembered that DF-HTA is not intended to provide a one-off assessment of a binary decision but to inform developers on an ongoing basis (see Table 4-1). Developers need to weigh a range of different

considerations in their decisions about how and whether to develop the technology, the role of DF-HTA is to provide input into these decisions.

5.3.3.3 Literature review

Literature review is a key method in DF-HTA where evidence is limited. Information from the literature such as national and local guidelines can inform clinical pathway analysis thus providing the decision model which is the basis for the economic model. Literature review may identify current and potential comparators (effectively a form of horizon scanning). A literature review may also directly inform the economic model, providing parameter estimates for clinical effectiveness and costs.

Systematic literature reviews may be appropriate for some forms of DF-HTA where time and resources are not limited. Even the briefest systematic review using automated tools has been estimated to require over 60 man hours (Clark et al, 2020) and a typical review takes over 800 man hours estimate (Ba' Pham et al, 2018). In situations where resource and time are constrained which is likely to be the case for much DF-HTA, adaptations to literature review methods are required. Abel et al (2019) recommend a 'targeted' approach where searches are conducted on a flexible, iterative basis and are interspersed with expert elicitation and stakeholder consultation. Clinical experts may be able to identify key literature sources such as relevant recent trials, guidelines or comprehensive HTA reports. Chapman (2013) recommends using a limited number of authoritative sources. Others have relied on particularly relevant single sources (Dranitsaris et al, 2004; Brandes et al, 2015). In all cases where parameter estimates were taken from a limited range of sources, the estimates should be confirmed with clinical experts and varied widely in sensitivity analysis (Grutters et al. 2019).

5.3.3.4 Multi-criteria decision analysis

Multi-criteria decision analysis was defined by Ijzerman and Steuten (2011) as 'a series of techniques aimed at supporting decision-makers faced with making numerous and sometimes conflicting evaluations. Multi-criteria decision analysis aims at highlighting these conflicts and deriving a way to come to a compromise

in a transparent process.' There are different variations of the method including the 'budget pie' method (Ryan et al, 2001), Simple Multi-Attribute Rating Technique (SMART) and Analytic Hierarchy Process (AHP)(Goodwin and Wright, 2014). They all follow the same basic structure set out in Table 5-6 but differ in the mathematical sophistication of the weighting process. The 'budget pie' method is the simplest in that it asks participants to allocate a set number of points over the attributes according to their importance to determine weights. SMART uses swing weights to elicit weights (Goodwin and Wright, 2014). This involves asking respondents to compare a change from the least preferred to the most preferred value on one attribute with a similar change (swing weight) in another attribute. AHP involves identifying top level criteria and grouping other criteria under these (Wissing, 2012). Weighting and scoring are determined using pairwise comparisons of sets of two (sub) criteria. The relative importance of one criterion in comparison to the other is appointed on a nine-point ordinal scale in which one reflects equal importance or preference and nine extremely higher importance or preference. These varieties of MCDA vary in complexity and transparency with budget pie being the least complex and most transparent. Steps one to four in Table 5-6 require informants who may be a combination of the developers themselves, clinicians and patients and a difficulty of any form of MCDA is in choosing or having access to informants who accurately represent the ultimate decision makers. Methods of eliciting the decision context and attributes draw on the qualitative methods previously described.

Table 5-6: Steps involved in multi-criteria decision analysis

Step	Description
1	Set out the decision context and the alternatives (i.e. the technology in development and current standard/s of care)
2	Identify the relevant attributes (i.e. the characteristics of the technology and comparators which are relevant to the decision)
3	Weight the attributes (i.e. the relative importance of the attributes)
4	Score the alternatives (How well do the new technology and the comparators perform on each of the attributes?)
5	Rank the alternatives using the weighted scores

Examples of the use of multi-criteria decision analysis

Wissing (2012) used AHP as part of a mixed methods study described as a 'clinical case scenario analysis'. This study assessed the potential clinical value

of a bioartificial pancreas (BAP) compared to conventional islet and pancreas transplantation in patients with type 1 diabetes mellitus. Literature searches, semi-structured interviews and focus groups were used to complete steps one and two set out in Table 5-6. Relevant attributes were: the site of BAP implant; the type of intervention necessary for placement of the BAP; the amount of donor material required; and the dose of immunosuppressive agents necessary for a successful transplant. In this example, three scenarios were constructed for BAP, a positive, a negative and a likely scenario in order that the preferences for the new technology over the current alternatives were assessed at different performance levels. To develop the weights, 12 endocrinologists and 27 patients completed a questionnaire. The three scenarios for BAP and the current alternative methods were scored by a clinician familiar with all the methods then these scores were weighted using the results from the AHP. The most important attributes were effectiveness, patient safety and technique for placement. Endocrinologists rated patient safety first whereas patient ranked effectiveness ahead of safety. The new technology was preferred by endocrinologists providing it reached most likely performance. Patients preferred conventional pancreas transplant in all performance scenarios.

Hummel et al (2012) assessed a new treatment for adolescent idiopathic scoliosis (AIS). The current standard of care involved bracing then posterior fusion surgery for severe cases. Posterior fusion surgery stops growth in adolescents as vertebrae are fused and there is a relatively high rate of complications including pain and post-surgery infections. The new treatment, non-fusion surgery (NFS), would allow surgery at an earlier age as there would be no restriction of growth and may prevent progression of disease. There was no clinical effectiveness data for the new treatment. A panel of four biomedical engineers and two orthopaedic surgeons estimated the missing data on a decision tree comparing new and existing technologies on costs (materials and treatment) and effectiveness (quality of life, pain, back function, self-esteem, medical and technical complications). In this study, the decision tree provided no absolute estimate of clinical effectiveness, utilities or costs but the branch with the highest relative priority indicated the best decision alternative. The study found that the new technology was the preferred option for patients with severe AIS because of its impact on spine function and patient esteem. Further work was

required to reduce cost and improve prediction of those patients likely to progress to severe disease to improve clinical acceptance of the method. Assuming data was available for the existing treatment absolute values may have been calculable.

Koning (2012) used SMART to assess the potential for a POC test (µDtect, an electronic nose) for the diagnosis of infectious disease. Interviews with a multi-disciplinary group of informants identified relevant attributes for the comparison of the new technology and standard of care. The most promising indication had been previously identified in some qualitative work. 13 attributes were identified across three main categories – cost (capital cost, implementation cost, initial training cost , maintenance cost, cost of consumables , ongoing training cost, cost of labour ongoing), clinical performance (accuracy, time to identification, chance of contamination) and impact on workflow (ease of use, integration with other systems). Clinical performance received the highest weighting at 0.56 with accuracy comprising 0.274 of that weighting. Cost was weighted 0.2 and impact on workflow 0.24. µDtect scored well in both cost and impact on workflow categories but it was not sufficiently accurate or rapid to offer advantages in well-resourced settings.

More examples of MCDA from applied studies identified in the review are provided in Appendix 9.

Strengths of multi-criteria decision analysis

- MCDA extends the qualitative clinical case assessment providing some quantification of informants' preferences for the new technology compared to the existing alternatives.

- MCDA methodology may encourage more explicit consideration of all relevant criteria. For example, if MCDA is used as an alternative to cost-effectiveness analysis to perform an economic evaluation, more criteria can be explicitly considered in the decision-making process and cost can be weighted appropriately for the decision context.

- The ability to assess different scenarios provides an indication of the importance of achieving certain performance criteria across different stakeholder groups.

Limitations of multi-criteria decision analysis

- The initial steps of MCDA involve the same narrowing down of the potential indication for a new technology as qualitative methods of stakeholder interaction. A negative result does not, therefore, indicate that the development of the technology should be stopped but may indicate that a new indication, application or setting may be more promising for the technology.

- Undertaking the simplest form of MCDA (e.g. budget pie) is not technically challenging and may be achieved without significant commitment of resource. Other forms of MCDA may require the use of specialist software or access to expertise that is not available for many developers of medical technologies at early stages of development.

- The quantification of preferences using methodology such as SMART and AHP may result in a lack of transparency. Moreover, the presentation of detailed quantitative results may be a case of 'false precision' (Pirie, 2015) given the lack of evidence early in development and a potentially limited range of settings and informants.

- Informants may not be representative of the real decision-makers.

5.3.3.5 Discrete choice experiment

A discrete choice experiment (DCE) is 'a quantitative method increasingly used in healthcare to elicit preferences from participants (patients, payers, commissioners) without directly asking them to state their preferred options' (York Health Economics Consortium, 2016). Participants are presented with a choice between hypothetical scenarios which include a number of attributes thus indirectly revealing their preferences for each of the attributes.

Example of discrete choice experiment

Groosthius-Oudshoorn et al (2014) used a discrete-choice experiment to estimate the clinical value in terms of increased test uptake for a variety of methods of screening for colorectal cancer (CRC) including a potential nanopill with 2,225 respondents from the UK and Netherlands recruited by random sampling among a population 50-75 years of age. Characteristics were identified from previous preference studies in CRC. The six attributes included in the DCE were preparation, technique, sensitivity, specificity, complication rate and required frequency of testing. The DCE revealed that participants preferred no preparation, 100% sensitivity and the nanopill compared to other techniques. The study also found that switching to a nanopill would result in an increase of 3% in uptake of screening, from 76% to 79%.

Latimer et al (2011) is an example of a brief time-trade off exercise being used to estimate utilities for an economic model.

Strengths of discrete choice experiments

A well-conducted DCE can provide evidence to developers of whether a new technology has the potential to outperform the existing standard of care. It also provides evidence about the relative importance of the different attributes and information to inform modelling.

Limitations of discrete choice experiments

It requires expertise and resources to conduct a DCE and these may not be available to SME developers of medical technologies.

5.3.3.6 Expert opinion and expert elicitation

An expert is defined by Morgan (2014) as an individual 'whose knowledge can support informed judgement and prediction about the issues of interest'. Iglesias et al (2016) distinguish expert opinion and expert elicitation on the basis that expert opinion involves qualitative information and expert elicitation quantitative information. Although both are considered weak evidence (Evans et al, 2003), they are useful for DF-HTA as stronger evidence is often limited.

Expert opinion may be used to provide background information on the disease area, current and potential clinical pathways, to help interpret evidence, to define the decision problem (including relevant comparators and populations most likely to benefit from a new technology) and to sense check a model structure, data inputs and results (Peel et al, 2018). Expert elicitation uses quantitative methods of elicitation and estimates may include probability distributions or ranges (Girling et al, 2007). Such estimates may be used to populate decision models for cost-effectiveness analysis.

Methods for eliciting expert opinion include qualitative stakeholder interaction methods discussed above. A more formal process is a Delphi panel which includes two rounds of consultation with a view to establishing consensus. Best practices for the reporting of Delphi processes to collect expert opinion and structured expert elicitation are presented by Iglesias et al (2016). A recent methodological report put forward a framework for expert elicitation with 20 recommendations across aspects such as selection of experts, avoiding bias, weighting judgements and reporting (Bojke et al, 2019). The report included an empirical evaluation of this framework and found that a total of five months of full-time equivalent researcher time was required to undertake the elicitation exercise. Even in the relatively well-resourced environment of NICE, time and resource constraints often lead to expert elicitation being conducted informally with individual experts (Peel et al, 2018).

For well-resourced DF-HTA, quantitative expert elicitation may be possible using available tools such as the Sheffield Elicitation Framework (SHELF) (Oakley and O'Hagan, 2010) and MATCH uncertainty elicitation (Morris et al, 2014). A number of applied examples were identified from the literature and are set out in Appendix 8. Where resources were limited expert elicitation would need to use more informal methods.

Examples of expert elicitation

Two of the studies identified in the literature review applied expert elicitation techniques in clinical value assessment, Breteler (2012) and Haakma et al, (2014). Breteler (2012) used structured expert elicitation with 18 neurologists to narrow down the potential indications for a home-based brain monitoring system for patients with suspected epilepsy and to assess their probabilities for success. She used interviews with two neurologists to map the diagnostic pathway, identify uncertainties in the pathway and several places where the technology may impact. She then defined five potential scenarios for use over three different populations. She elicited the mode and range for a beta distribution of the probability that the neurologist would order the home-based monitoring at that particular point for a particular patient. She also included a weighting of clinician estimates according to their years of experience, the number of electroencephalograms done each week and their ability to make long term recordings. Haakma et al (2014) used structured expert elicitation to elicit radiologists' beliefs about a range of test performance parameters in order to assess the clinical value of a novel imaging technology. The study compared photo-acoustic mammography (PAM) to magnetic resonance imaging (MRI) in suspected breast cancer after an inconclusive xray or mammogram. 20 radiologists from academic and non-academic hospitals in Netherlands were asked to estimate the number of true positives and true negatives likely to result from examination of a range of images. First, the experts were asked to rate the importance of seven tumour characteristics in the examination of images by allocating 100 points across the characteristics. Then they were asked to grade each imaging modality with its ability to visualise those characteristics, again giving a score between 0 and 100. A questionnaire and spreadsheet-based exercise was used to elicit the probability distribution of true positives and true negatives during a face to face interview with individual radiologists. They estimated TP and TN compared to pooled data for MRI by estimating the mode and the 95% credible intervals with the support of a graphical display of the probability density function.

Structured expert elicitation can also provide quantitative estimates of parameters for use in economic evaluations before clinical evidence is available.

Probability distributions can be elicited as well as point estimates, which can enable the development of models using PSA and VOI. Kip et al (2016) elicited the effect of a new biomarker on hospital discharge rates and interventions performed. The elicitation involved the administration of a questionnaire to ten cardiologists. The discharge and intervention rates provided parameter estimates for a cost-effectiveness model.

Strengths of expert elicitation

- The use of expert judgement may be the only way to source information, whether qualitative or quantitative, about the likely performance of a new technology. However, there is concern that experts are often over-optimistic about the likely performance of a technology and underestimate uncertainty, particularly if the experts are also involved in the development of the technology (Gosling, 2014). More structured methods of expert elicitation try to ensure that heterogeneity and uncertainty are fully represented (Gosling, 2014).

- Expert elicitation of probability distributions as well as point estimates enables the use of probabilistic analysis in models.

Limitations of expert elicitation

- Structured expert elicitation requires some resource and expertise to undertake and contacts with experts which may be difficult for SME developers. The estimation aspects of structured expert elicitation may sometimes be difficult for experts to operationalise. For example, six out of the 20 radiologists who undertook the initial elicitation exercise in Haakma et al (2014) felt unable to complete the elicitation of true positives and negatives.

5.3.3.7 Cost-effectiveness analysis

As shown in Figure 5-1, in the generic process of DF-HTA, economic value assessment follows clinical value assessment. It accounts for healthcare and other resource usage. It may also include productivity and other wider impacts.

In reimbursement-focused HTA, economic evaluation is typically undertaken to determine whether the technology, at a given price and given current evidence, represents a cost-effective use of healthcare resources. In development-focused HTA, economic evaluation using health economic modelling is also the mainstay, but it plays a different role. It may be used to support a range of decisions including whether further development should be funded (go/no-go decision), placement of the technology (setting, indication and position in clinical care pathway), and threshold performances required or prices achievable for the technology.

Economic value assessment in DF-HTA may include a formal cost-effectiveness analysis. The umbrella term, cost-effectiveness analysis, includes cost-benefit, cost-utility, cost-effectiveness, cost-consequence and cost-minimisation analysis (Drummond et al, 2015). All these methods compare the difference in outcomes brought by a new technology to the difference in costs. All methods calculate costs in the same way, but they measure outcomes differently. In a cost-benefit analysis, outcomes are measured in monetary terms. In a cost-utility analysis, outcomes are expressed using a generic measure such as quality adjusted life years (QALYs). This is because the generic outcome measures used in cost-utility analysis facilitate comparisons across technologies and disease areas. In cost-effectiveness analysis, outcomes are expressed as disease-specific measures, for example, cost per infection avoided. In cost-consequence analysis, outcomes are expressed across multiple measures such as cost and cancer cases missed. A particular challenge with both cost-effectiveness and cost-consequence analysis is the difficulty in making comparisons between indications and determining an acceptable threshold for willingness to pay. Finally, in cost-minimisation analysis it is assumed that outcomes are either equal or superior for the new technology, hence it is sufficient to simply compare costs. The challenge with this analysis is that it is only appropriate when it is safe to assume that outcomes are either equal or superior with the new technology.

In some jurisdictions (for example: Australia, Canada, Netherlands, UK) there is a strong emphasis on cost-effectiveness analysis in reimbursement programmes and methods are well established (PBAC, 2016; CADTH, 2017; National Health Care Institute, Netherlands. 2016; NICE, 2013). Where the target audience for

DF-HTA is not a publicly accountable body (such as a reimbursement agency) there is not the requirement for methods to be uniform and pre-prescribed. Commercial developers are investing their own (or their shareholders') money and ultimately their success will be measured on whether the investment proved worthwhile rather than the quality and transparency of the process supporting the decision. Commercial developers are more likely to take a pragmatic view of the methods used as they need to balance the expenditure of effort and funds between numerous aspects of the development. This means that analysts of DF-HTA can also take a more pragmatic stance of analysis as the methods to be used are not standardised. They can apply methods which are appropriate for the developers' needs and meet a burden of proof acceptable to them. Public sector funders of research may require more structured methods to be applied as they may have to demonstrate consistency and transparency of their decision-making process.

As a result of these features (see also Section 4.2.2.1) some adaptations to these established methods might be expected when the analysis is being undertaken as part of DF-HTA. A suggestion of such adaptations is set out in

Table 5-7 and discussed in the paragraphs that follow. The comparison is made against the reference cases put forward by the technology appraisal and diagnostic appraisal programmes (TAP and DAP) from NICE in the UK (NICE, 2013).

Table 5-7: Adaptations in economic evaluation for development-focused HTA

Aspects of economic evaluation	Economic evaluation for coverage decisions (based on NICE TAP and DAP methods guides)	Hypothesised adaptations for development-focused economic evaluations
Timing	Market access stage	Pre and during development Ideally an iterative approach Less evidence More uncertainty about positioning and performance of the technology
Source of decision problem	NICE scoping exercise	Context dependent Expert opinion Less rigorous methodology
Perspective	UK NHS and PSS	Multiple markets including public and/or private payers
Type of economic evaluation	Cost-utility analysis for TAP More flexible for DAP	Focus on cost minimisation/ health outcomes other than QALYs Less focus on CUA
Time horizon	As required for health and cost outcome differences	Shorter time horizons – may not capture all differences
Discount rate applied	3.5% discount on costs and health outcomes	Multiple discount rates Discount rates not applied
Evidence for health effects	Systematic review. RCT preferred for TAP More flexible for DAP	Literature review not systematic Small clinical studies Expert opinion or assumptions
Evidence of quality of life	Health state reported by patients, valued by UK population	Literature review not systematic Expert opinion or assumptions
Evidence for costs	Systematic review for TAP More flexible for DAP	Comparator information Expert opinion or assumptions
Modelling method	Unspecified but should be fit for purpose	Simpler modelling methods (headroom calculations, decision trees, markov models)
Estimand	ICER, NMB Budget impact	Wider range including headroom and threshold costs, performance, health outcomes. Budget impact.
Dealing with uncertainty	Probabilistic sensitivity analysis for parameter uncertainty Sensitivity analysis including alternative data sources Scenario analysis for structural uncertainty	Multiple scenarios to reflect uncertainty about positioning of technology One-way sensitivity analysis for parameter uncertainty in earliest stages with probabilistic sensitivity analysis when some clinical evidence is available.

CUA – cost-utility analysis, DAP – Diagnostic Assessment Programme, EE – Economic evaluation, HTA – Health Technology Assessment, ICER – Incremental Cost Effectiveness Ratio, N/A – not applicable, NHS – National Health Service, NICE – National Institute for Health and Care Excellence, NMB – Net monetary benefit, PSS – Personal Social Services, QALY – Quality Adjusted Life Year, RCT – Randomized controlled trial, TAP – Technology Assessment Programme, UK – United Kingdom

1. Timing

NICE assessments are typically done at the market access stage. Economic evaluation as part of DF-HTA is generally undertaken before and/or during the development of a technology on an iterative basis although it can continue throughout the lifecycle. In the 32 examples of economic evaluation for developers described by Grutters et al (2019), one was an idea not yet being developed, 16 were in development, eight had finished development and seven were on the market. Although an iterative approach is recommended in the

literature (Grutters et al, 2019; Drummond, 2019) only one of the examples identified in the literature review presented an iterative approach and this was reconstructed on a retrospective basis as an illustration (Vallejo-Torres et al, 2011).

As economic evaluation in DF-HTA is often carried out before there is extensive clinical evidence specific to the technology. This impacts on the source of estimates for populating models. Uncertainty is greatest at the earliest stage in product development when the best placement for the technology may not yet have been determined. At this earliest time, it may be appropriate to undertake a less resource intensive form of economic evaluation alongside the articulation of the initial value proposition for the technology (Graziadio et al, 2020). Later in the development process, when the positioning of the technology is likely to be more concrete and there may be some clinical evidence specific to the technology a more developed form of cost-effectiveness may be appropriate (Graziadio et al, 2020).

2. *Source of decision problem*

For NICE assessments the scope of the decision problem is a collaborative, systematic and resource-intensive process (NICE, 2013, 2020a). For economic evaluation in DF-HTA, the initial decision problem may be defined in consultation with the developers, depending on where they see the technology being used and what advantages it is thought to have over the current standard of care (Grutters et al, 2019). Interviews with clinical experts may then be used to determine the current, and alternative, clinical pathways in sufficient detail to permit modelling (Grutters et al, 2019). In order to keep the modelling exercise manageable, expert input may also be required to narrow down the alternative strategies to the most plausible ones although this approach is not always taken (Faria et al, 2018).

3. *Perspective*

A UK national health service and personal social services (NHS and PSS) perspective is required by NICE guidance. This means that, generally, the only costs included are those incurred by the NHS or for adult social care. Costs

incurred by the patient and productivity losses are excluded. In economic evaluation as part of DF-HTA, multiple markets may be considered or a wider consideration of costs and benefits may be appropriate in a study, as technologies may be offered to private payers as well as to healthcare systems or to healthcare systems which do not place the same emphasis on cost-effectiveness analysis as the UK. In the applied studies identified in the review there was little evidence of the consideration of multiple jurisdictions (see Appendix 7).

4. Type of economic evaluation

NICE requires a cost utility analysis to be undertaken and may also require a budget impact analysis. Cost-utility analysis uses the generic metric, the quality adjusted life year (QALY) as health outcome. This allows NICE to compare across different disease categories on an equivalent basis. A budget impact analysis is required when the anticipated budget impact is high. In economic evaluation as part of DF-HTA, outcomes other than the QALY may be used as developers seek to demonstrate wider views of value than those captured in a QALY or lack the evidence translating clinical effectiveness into QALY values. There may also be a focus on cost saving or overall budget impact as this may be the easiest way for developers to demonstrate the value of their technology to a decision-maker.

A type of cost-utility analysis suited to DF-HTA because of its simplicity is what has been termed 'headroom analysis' (Ijzerman and Steuten, 2011; Chapman, 2013; and Girling et al, 2015). In fact, headroom is simply an estimand from any form of cost-effectiveness analysis. 'Headroom' is an indication of the maximum price which could be charged for a new technology. It is calculated by multiplying the increase in utility (or other relevant effectiveness metric) expected to be generated by the introduction of the new technology by a willingness to pay threshold (for the relevant metric) for a given jurisdiction then adjusting for cost savings/additional costs included in the calculation. In the UK, there is an explicit willingness to pay threshold of £20,000 to £30,000 per QALY (NICE, 2013). For example, if a new technology was expected to generate an incremental QALY of 0.1 and save £500 in further diagnostic procedures the headroom would be calculated as 0.1 x £20,000 (representing the

willingness to pay for 1 QALY in the UK) + £500. A total headroom of £2,500. The headroom estimate represents the maximum price achievable for the technology if it is to be considered cost-effective in the UK. Threshold analysis is a similar formulaic approach, which assumes a price for the technology then investigates what values the other parameters need to take in order for the technology to remain cost-effective. Both headroom and threshold analysis are simple and quick to perform and can be based on expert opinions or assumptions, so they are ideal for undertaking extensive scenario or sensitivity analysis and when resources are constrained (Chapman, 2013; Vallejo-Torres et al, 2008). The analysis may focus only on cost savings rather than valuing health outcomes (Markiewicz et al, 2016) and may even be limited to the analysis to a single category of savings (McAteer et al, 2007). Craven et al (2009) developed the MATCH headroom tool, available on-line (Match.ac.uk, 2018), which can be used to undertake a headroom analysis.

Cost-minimisation (Frappier et al, 2014; Steuten et al, 2009) or cost-consequence analysis (Brandes et al, 2015) should theoretically be useful where resources are constrained although not many examples exist in the literature (Hunter and Shearer, undated). The methods are useful because they avoid the evaluation of health outcomes in terms of utilities (required for cost-utility analysis) which may be complex. Cost-effectiveness analysis where the outcome of interest is a clinical measure, may also be useful in DF-HTA as it avoids the use of a utility measure like the QALY. Applied studies were identified in the review which used outcomes including complications avoided, cancer cases prevented, correct diagnoses, time to treatment and patients correctly discharged. Budget-impact analysis goes beyond cost-consequences analysis in that it takes account of the scale and timing of revenues and expenditure associated with the new technology. It is particularly important to make budget impact transparent where spending in one area is offset by savings in another (Graziadio et al, 2020).

5. *Time horizon used*

The time horizon required by the NICE methods guides is a period sufficient to allow all differences between health outcomes and costs for the alternative technologies to be fully explored. For many technologies this is equivalent to a

lifetime horizon. As modelling to a lifetime horizon is often technically challenging and resource intensive, economic evaluations in DF-HTA may use shorter time horizons particularly in the earliest stages of development (Brandes et al, 2015; Chapman, 2013).

6. Approach to discounting

NICE requires that both costs and health outcomes are discounted by 3.5% per annum. In economic evaluation for DF-HTA, due to a desire to simplify analysis and shorter time horizons adopted, analysts may not apply a discount rate.

7. Evidence for health effects

NICE requires evidence of health effects to be derived from a synthesis of the evidence from the literature extracted by systematic review. In the TAP programme, evidence would generally from randomised controlled trials. The DAP programme is more flexible and will accept other forms of high-quality evidence specific to the technologies under evaluation. Due to the timing of economic evaluation in DF-HTA, robust clinical evidence specific to the technology may not be available. Alternative sources of parameter estimates may be literature (see Section 5.3.3.3), outputs from clinical trial simulation, expert opinion or assumptions. The analyst may also choose to make assumptions. This is often done for the price of a technology at concept stage. In all these cases estimates would be varied over a wide range in sensitivity analysis and limitations of the data reported transparently (Chapman, 2013; Vallejo-Torres, 2011).

8. Evidence for quality of life

For quality of life estimates, NICE requires patient reported health conditions to be valued by the UK population using preference-based methodology. In economic evaluation for DF-HTA, directly obtained estimates are often not available and estimates may be required for more than one jurisdiction. Quality of life estimates may be derived from the literature in a non-systematic way or provided by experts. Chapman (2013) described an approach for adapting utilities from the literature termed the ‘utility ladder’. This approach (based on

solutions adopted when few utility studies were available) (McAteer, 2011) involved deriving utilities by placing the unknown condition among a range of conditions for which utilities were available. Other analysts have used simple elicitation exercises to estimate utilities (Dranitsaris et al, 2004; Latimer et al, 2011; McAteer et al, 2007; Vallejo-Torres et al, 2011).

9. Evidence for costs

In the technology appraisal programme, NICE requires that evidence for costs is derived using systematic review methodology although this is a little more flexible in the DAP. For economic evaluation in DF-HTA, similar to the position with other categories of evidence, resource constraints and the timing of the analysis means that evidence on costs is more likely to come from less rigorous literature searches, relate to comparator technologies or be based on expert opinion or micro-costing exercises. Different cost levels may be assumed or the analysis may work backwards to determine what the threshold cost may be for the technology to be considered cost-effective in a particular jurisdiction.

10. Modelling method used

NICE do not specify which modelling method is preferred providing that it is fit for purpose. Decision trees, Markov models and simulation models are all regularly used. A decision tree is a 'visual representation of all the possible options and the consequences that may follow each option' (Hunink et al, 2014). The initial line of a decision tree is labelled with the population or problem you are considering then a square indicates a decision node at which one of the alternative options, represented by subsequent lines, must be chosen (Hunink et al, 2014). The circles (or chance nodes) represent time points at which there are two or more possible consequences of a decision. These consequences are assigned probabilities which must sum to 1. If cost and health outcomes are attached to each branch of the tree, the overall expected values for costs and outcomes can be calculated for each strategy.

One practical difficulty with decision trees is that they can become unwieldy quickly particularly if longer time horizons need to be considered and/or if the disease is of a remitting/relapsing type (Briggs et al, 2006). As a consequence,

decision trees generally represent a relatively short time horizon, often the acute treatment phase of an intervention. The Markov framework is often used to extend the short time horizon covered by a decision tree. Markov models are structured around mutually exclusive disease states. Rather than representing different possibilities over time as pathways in a decision-tree they are modelled as possibilities of transitioning between different disease states during a discrete time period known as a cycle (Briggs et al, 2006). Costs and health outcomes are attached to each of the disease states and they accumulate based on the number of individuals from a notional cohort in any one state during a given period. Markov models often include a death disease state and include sufficient time periods to allow the analysis to extend until all members of the modelled cohort have transitioned to the death state.

A significant limitation of Markov models is that they are said to be 'memoryless' (Briggs et al, 2006). This means that the model type, in its simplest form does not take account of how long an individual in a particular disease state has been in that state. This can limit the ability of the Markov model to accurately represent some disease processes. For example, in many types of cancer the probability of progressing to more severe forms of the disease may depend upon how long the patient has been in remission following the first treatment. This aspect would be difficult to capture in a simpler Markov model. Although adjustments to the Markov model have been put forward, they may result in unwieldy models which are difficult to operationalise (Briggs et al, 2006).

Simulation models, where patients pass through the stages of the model on an individual basis, are a possible solution to this limitation. They allow the accumulating patient history to impact upon transition probabilities, costs and health outcomes (Briggs et al, 2006). Simulation models may require more data, which are generally not available in development-focused HTA. However, if a large dataset (of observational data, for example) were available, this could be developed into a generic decision model which could then be available to address a range of decision problems in the disease area as they emerge (Briggs et al, 2006).

Conceptual models are a means of representing an understanding of the nature of the problem under consideration, the objective of a decision model and its

scope prior to the development of the decision model (Roberts et al, 2012). They generally precede a quantitative decision model although in theory the elucidation of the conceptual model, in itself, may be sufficient to inform a negative go/no go decision where there is no clear mechanism for the new technology to deliver incremental benefit or cost savings compared to the current standard of care.

In economic evaluation for DF-HTA, simpler models may be preferred as they are easier to construct and are more transparent (Grutters et al, 2019). For example, Craven and Morgan (2011) were able to develop a simple Markov model in two weeks. Headroom and decision trees are easy to understand and thus good to aid communication between the analyst and non-expert. Complex methods used in the early stages of development when there is uncertainty over the positioning of the technology and no clinical evidence may risk false precision (Grutters et al, 2019). Simple models are not universally preferred. Annemans et al (2000) argued that it was more useful and efficient to produce a model of the same quality required at reimbursement stage earlier in development. This ensured that the full complexity of the clinical and disease pathways were captured and identified all evidence gaps to aid evidence generation strategy. This point of view is also taken by two recent studies setting out the experience of two of the Diagnostic Evidence Cooperatives, public-funded bodies in the UK, set up with aim of helping developers of diagnostic tests to generate evidence (Abel et al, 2019; Graziadio et al, 2020). Abel et al (2019) recommend the development of flexible models which can be updated iteratively as new clinical evidence becomes available. The authors do recognise that such models are not necessarily faster, easier to implement or less complex than models developed at late stages of development so that resourcing is a challenge, particularly if more than one scenario needs to be modelled.

11. Estimand

NICE generally require the estimation of an incremental cost-effectiveness ratio (ICER) or Net Monetary Benefit (NMB) and budget impact where this is likely to be over £20 million per annum (NICE, 2020b). Broader considerations such as equity are considered in decision-making committees. In many jurisdictions

cost-effectiveness is not considered as part of the choice of treatments and even in jurisdictions where it is considered, it is not used in all situations. In economic evaluation for DF-HTA, both developers and decision-makers may be interested in wider considerations than those captured in an ICER. Considerations for developers include whether to continue to invest, where to position the new technology, how to price it, what the performance of the technology needs to be to justify a given price and what the most important drivers of cost-effectiveness may be to inform evidence generation strategies. Both developers and decision-makers may also be interested in implementation issues such as budget impact including timing of cash flows, how well the technology fits into the existing clinical pathways and workforce configurations.

An ICER or NMB can be estimated in DF-HTA. Grutters et al (2019) estimated an ICER when there was data for both costs and effects. This is useful to developers as an indication that, on the basis of existing evidence, the technology would be considered cost-effective in a given jurisdiction and may be recommended or reimbursed in that jurisdiction. All 30 technologies considered by Grutters et al (2019) could potentially have been cost-effective using the threshold willingness to pay in the Netherlands. This led Love-Koh (2019) to comment that potential cost-effectiveness is a 'low bar' and not a very useful finding for developers. However, Grutters et al (2019) make the point that their sample may be biased towards technologies with more potential as the developers have had the confidence to commission a health economic model. Two of the studies identified in the literature review had negative findings (Markiewicz, 2017b; Kluytmans et al, 2019) which would seem to indicate that health economic modelling can be useful to inform decisions about whether to continue to invest in the technology. Negative findings in the indication, setting or position in a clinical pathway suggested by the developer may also help to inform decisions about where to position a technology. For example, Kluytmans et al (2019) found that a surgical instrument believed to be of benefit in meniscus surgery would not add much value in that indication but there was potential in other surgical applications.

Applied studies identified in the literature review estimated the 'headroom' of a technology in development (for example Chapman, 2013; Markiewicz et al, 2016;

Vilsboll et al, 2018). Headroom indicates the maximum price which can be charged for a technology at given performance levels assuming the developer 'captures' the full added value of the technology. This can be useful to developers if the headroom is much lower than their price expectations they have the ability to reconsider the chosen market, change the design of the technology (for example, aiming to improve test performance) (Drummond, 2019). The price estimate may also be used as an input to a return on investment (ROI) calculation where the developer compares the likely discounted cash flows generated by the new technology to the future costs of development in order to estimate the likely profit arising (Singh, 2010; Markiewicz et al, 2016). Threshold technology performance may be estimated if a price or set of prices are assumed (Grutters et al, 2019). In diagnostics, for example, modelling a number of scenarios with the new technology used in different positions in the clinical pathway may help to identify the threshold sensitivity and specificity required for the new test to be considered cost-effective in the given jurisdiction (Miquel-Cases et al, 2017).

Although not a quantitative estimand, Grutters et al (2019) included aspects relating to the implementation of the technology to the developers who had commissioned the health economic modelling. Lehoux and Silva (2019) suggest that 'risks, competing upcoming innovations and logistical issues' be more explicitly considered in order to reduce failure rates and inefficiency in research and development. Although implementation and adoption may be considered beyond the scope of DF-HTA, the modelling process may make evident some barriers to or facilitators of adoption that are of interest to developers in making decisions during the development process.

12. Methods of dealing with uncertainty

The NICE TAP and DAP methods guides require different approaches to deal with the various sources of uncertainty inherent in economic evaluation. For structural uncertainty and uncertainty around which data source to use, sensitivity analysis is recommended (NICE, 2013, 2020a). For parameter uncertainty, probabilistic sensitivity analysis (PSA) is recommended. PSA deals with uncertainty in all parameters simultaneously compared to one-way

sensitivity analysis, which varies one parameter at a time whilst holding the other parameters constant.

In addition to the three forms of uncertainty referred to above economic evaluation for DF-HTA may have a further source of uncertainty due to the wide decision space available to the developers. Until the positioning of the technology is decided there may be many combinations of setting, indication and position in clinical pathway which require investigation. Parameter uncertainty is also high in DF-HTA as there is often no clinical evidence specific to the technology and little information on how the technology may be priced. Grutters et al (2019) reported the use of deterministic models and one-way sensitivity analysis (OWSA) because they are simpler, less resource intensive and are easier for non-experts to understand. Deterministic models are those which assume a single value for a parameter. OWSA allows decision-makers to understand the impact of individual parameters on the results of the analysis, particularly when presented visually, for example using a tornado diagram. Grutters et al (2019) were sceptical of PSA as they felt it may give an impression of 'pseudo-certainty' when there is so little data. The use of simpler models allows the exploration of multiple scenarios at a time when resources may be scarce. Scenario building is an approach originating in management science (Retel et al, 2012). It is similar to sensitivity analysis in that input parameters are varied in order to determine the impact on outcomes of interest. However, it involves the imagining of a number of potential future scenarios, for example, concerning the future diffusion of a health technology or its relative performance compared to current standard treatment. The parameters are varied in accordance with the imagined scenario, resulting in a number of parameters being varied at one time.

The approach described by Grutters et al (2019) and reflected in some of the studies identified in the literature review (for example, Craven et al, 2011; Hummel et al, 2012; Latimer et al, 2011) of simple deterministic models and OWSA has been criticised by some commentators (Love-Koh, 2019; Abel et al, 2019; Federici and Torbica, 2020; Kim et al, 2020). Love-Koh (2019) commented that the methods adopted 'fall short of the methodological standard of probabilistic sensitivity analysis and value of information uniformly

recommended in the literature for early models'. In support of this argument, he cites Annemans et al (2000) who argue for models alongside pharmaceutical development to include all the complexity of the final reimbursement focused model so that the model can be used throughout the life-cycle of the drug. Graziadio et al (2020) also support this line of argument claiming that it is more efficient to create a model which is sufficiently flexible to allow iterative updating as more information becomes available. They do, however, recognise that this may be challenging due to restrictions on resource. Vallejo-Torres et al (2008) suggest a two-stage process where at the earliest stage simple models and OWSA are used but when some clinical evidence is available a more complex model is built including PSA to allow value of information analysis. As simple methods do not require much resource and many technologies will fail, this approach may be equally efficient to producing a complex, flexible model at an early stage.

Strengths of cost-effectiveness analysis

- Simple, deterministic models provide a low resource method to explore potential technology positioning across wide decision space. Such models are useful as aids in communication, to guide developers about price and technology performance thresholds and to identify parameters likely to impact most on cost-effectiveness.

- Once the positioning of the technology is more certain and there is some clinical evidence, more complex models, potentially incorporating probabilistic sensitivity analysis and value of information analysis, may be developed which would help inform evidence generation strategies. Such a model could be updated on an iterative basis as more evidence became available and be used to support early engagement with a reimbursement agency, if that was relevant.

Limitations of cost-effectiveness analysis

- Cost-effectiveness analysis has been described as reductive (Grutters et al, 2019; Lehoux and Silva, 2019) in the sense that it only considers the cost and clinical effectiveness of a technology when other issues are of

importance to decision makers. However, although the quantitative analysis may be reductive, the process of constructing the model may provide insights into aspects which may help or hinder the technology and these can be communicated qualitatively to the developers alongside the results of the cost-effectiveness analysis (Chapman, 2013; Grutters et al, 2019).

- More developed forms of cost-effectiveness analysis may be resource and expertise intensive and is competing for funds at a time when many developers do not have unlimited resource. Cost-effectiveness analysis itself must deliver value for money for the developer. This value could be in terms of improving the chances of success of the technology, perhaps by demonstrating the levels of test performance required in a given position. The analysis may reduce overall development costs if technologies with little chance of success were identified at an early stage of development.

5.3.3.8 Value of information analysis

If probabilistic sensitivity analysis is undertaken alongside cost-effectiveness analysis the characterisation of parameter uncertainty can be used to provide estimates of the potential value of additional information from further research. Value of Information (VOI) analysis can be conducted from a commercial or a societal perspective. When conducted from a societal perspective the analysis can be used to estimate the net impact on health benefit arising from an increase in the probability of selecting the optimal treatment resulting from further research reducing uncertainty. Expected value of perfect information (EVPI) is the basic form of VOI and it values the resolution of uncertainty across all parameters simultaneously. A development of EVPI is expected value of partial perfect information (EVPPI) which values reduction in uncertainty of individual or groups of parameters. This is useful to identify the parameters which most influence the cost-effectiveness estimate so should be the focus of evidence generating activities. Expected value of sample information (EVSI) values the reduction in uncertainty to be expected from a specific trial with a given sample size. Expected net benefit of sampling (ENBS) is the difference between EVSI and the cost of the trial (Steuten et al, 2013). Undertaken from a

commercial perspective, the value of the reduction in uncertainty can inform pricing decisions and allow developers to balance the cost of a future clinical trial to the improvement in prices which may be available if there is greater certainty of cost-effectiveness (Drummond, 2019).

VOI uses the uncertainty estimates from PSA and may be difficult to operationalise and of questionable worth when there is limited data specific to a technology. However, as the evidence base starts to develop, VOI may be useful to determine whether further research is worthwhile and if so, what the optimum design for a trial would be. Vallejo-Torres et al (2011) illustrated how VOI can aid study design and investment decisions for commercial technology developers. Both expected value of perfect information (EVPI) and expected value of perfect parameter information (EVPPI) were estimated in the study. VOI may be particularly useful in situations where a technology is approved with further research and research is being funded by a public payer. For example, Miquel-Cases et al (2016b) used EVPI, EVSI and ENBS to inform decisions about the design of future studies and to prioritise research and Boyd (2012) used EVSI to determine appropriate sample sizes for future research studies.

There is some overlap between VOI and both clinical trial simulation (CTS) and assurance calculations. VOI builds on the output of cost effectiveness analysis to estimate the value of future research in terms of a reduction in uncertainty. CTS is a form of modelling and simulation used generally within pharmaceutical companies in order to model biological systems and the pharmacology of treatments acting on these systems (Holford et al, 2010). Assurance calculations are an alternative to statistical power calculations (O'Hagan et al, 2005). Both CTS and assurance calculations are techniques designed to inform the efficiency of clinical trials thus having a role in the research and development process. Neither method is concerned with the assessment of either clinical or cost effectiveness of a technology in development although both may use similar inputs and similar modelling techniques.

5.4 Summary and contributions of this chapter

This chapter set out a generic process for development-focused HTA (DF-HTA) comprising two linked activities, clinical and economic value assessment. It

distinguished these assessment activities from the linked activities of research and development and other commercial activities. Clinical and economic value assessment are both iterative and may be revisited a number of times during development. As much technology development could be described as 'technology-driven', assessment work in the earliest stages may involve 'positioning' the technology. At this point there may be no evidence specific to the technology available. Many methods share the approach of testing the developers' ideas about technology positioning with clinical experts who inform the analyst about the current clinical pathway and potential pathway if the new technology were adopted. The methods vary in terms of how they interact with the clinical experts and how they present their results. Qualitative methods of user interaction and headroom appear to be particularly appropriate at this earliest stage. Kluytmans et al (2019) provides a good example of the approach. This assessment in the earliest stage can confirm the potential (or otherwise) of the technology in the particular position, provide insights about potential barriers or facilitators to implementation and guidance on attributes of the technology which require modification in order to increase chances of success. The earliest stage of assessment can assist the developers to articulate the value proposition for the technology. As the development progresses, the positioning of the technology may be more certain and there may be some clinical evidence specific to the technology. At this point, developers and investors may be more confident of the potential of the technology and more willing to invest in more formal analyses. It may be appropriate at this stage to develop a cost-effectiveness model. Although there is some debate about whether simple or complex methods are appropriate, more formal methods of expert elicitation to populate the model are an option and these would allow the inclusion of probability distributions to enable probabilistic sensitivity analysis and value of information analysis. This form of model may form the basis of early dialogue with a reimbursement agency and/or a discussion around managed entry or risk sharing. Alternatively, if cost-effectiveness analysis is considered reductive, a formal multi-criteria decision analysis may provide evidence of the potential of the technology.

This chapter has two main contributions. Firstly, it presents a generic process for DF-HTA as a second element of a framework of DF-HTA. The generic process

builds upon the work of the MATCH UK collaboration and Professor Ijzerman's group including work undertaken at CTMM. Of relevance was the work of Cosh et al (2007), Davey et al (2011), Markiewicz et al (2014) and Rogowski et al (2016). The generic process comprises core assessments of clinical value and economic value which are linked to the parallel activities of business case development and research and development. The second contribution builds upon the work of Markiewicz et al (2014) to identify HTA methods used in each of these activity groups and to explain their appropriateness with reference to the features of development-focused HTA identified in chapter 4. This simplifies the existing literature where business case development and research and development were not clearly distinguished from assessment activities and methods of each had become conflated. It also distinguishes the earliest stages of assessment when less formal methods may be appropriate to confirm the positioning of the technology and later stages when more formal methods could be of use. This simplification and clarification is consistent with the recent addition to the literature from Graziadio et al (2020) and may be of assistance as a point of reference to analysts undertaking development-focused HTA.

6 Approach to case studies in development-focused HTA

6.1 Introduction

Chapter 6 provides a link between the methodological parts of the thesis and the empirical sections. It addresses research question 4:

4. To what extent do the case studies demonstrate the suggested features, process and methods of HTA to inform developers of health technologies?

Chapters 4 and 5 set out a framework of features, process and methods of development-focused HTA. After briefly introducing the case studies, this chapter draws on the outputs of the previous chapters and applies them to the case studies. First, the case studies are used to illustrate the list of features developed in Chapter 4 then activities and methods for the case studies are selected from those set out in Chapter 5.

The initial concept for this thesis was to undertake a prioritisation exercise in the context of the project base of Glasgow Molecular Pathology Node. For practical reasons this was not possible and the case studies eventually formed a convenience sample sourced partly from the GMP Node. Three case studies, all concerned with tests predicting treatment response were node projects (Case studies 2 and 3) or suggested by node participants (Case study 1). A further two case studies (4 and 5) were sourced through personal contacts and selected as they were useful illustrations of methods of development-focused HTA. A summary of the case studies is provided in Table 6-1. For the purposes of this thesis acute means ‘severe and sudden in onset’ and chronic means a condition which ‘develops over time’ (MedlinePlus, 2021).

Table 6-1: Summary of case studies included in thesis

Case study number	**1**	**2**	**3**	**4**	**5**
Chapter	**7**	**8**	**9**	**10**	**11**
Disease area	Ovarian cancer	Rheumatoid arthritis	Colorectal cancer	Malignant melanoma	Prostate cancer
Type of condition	Acute	Chronic	Acute	Acute	Acute
Technology type (see Table 2-1)	Prediction of response to treatment	Prediction of response to treatment	Prediction of response to treatment	Diagnosis in symptomatic or previously screened population	Diagnosis in symptomatic or previously screened population
Stage of development (see Figure 2-2)	Prior to basic research	Targeting specific biomarker	In clinical use	Targeting imaging modality	Available on the commercial market
Source of project	GMP Node investigator	GMP Node project	GMP Node project	Contact of supervisor	Contact of department

GMP – Glasgow Molecular Pathology

6.2 Illustration of features of DF-HTA

6.2.1 Introduction

Table 4-1 set out a list of ten features of DF-HTA and Table 4-2 set out questions which the analyst may wish to discuss with developers prior to undertaking an assessment. In this section, the features of each of the case studies are assessed against the features in the table in order to illustrate this element of the framework of DF-HTA. The questions are considered in relation to each case study. As the case studies were completed before the questions were answered, this exercise is less valuable than it may have been but may provide some insight into the usefulness of the questions. This section includes a table setting out the features of each case study. Section 6.2.2 to Section 6.2.7 discuss both the features and responses to questions for each case study in turn. Section 6.2.8 concludes on the extent to which the case studies demonstrate the features of DF-HTA and on the usefulness of the question

Table 6-2: Comparison of features of case studies with features of DF-HTA						
Feature of development-focused HTA	**Feature of Development-focused HTA**	**Case study 1** **Test to predict response to treatment in ovarian cancer**	**Case study 2** **Test to predict response to treatment in rheumatoid arthritis**	**Case study 3** **Extension of molecular testing in colorectal cancer**	**Case study 4** **Diagnostic test in malignant melanoma**	**Case study 5** **Diagnostic test in prostate cancer**
Target audience of	**Technology developers (both academic and commercial) Investors (both commercial and public sector)**	Academic and clinical development team Public funders	Academic and clinical development team Public funders	Clinical team Molecular Pathology Evaluation Panel	Small commercial technology company Commercial investor	Clinical team Hong Kong Hospital Authority
Underlying user objective	**Maximise long-term financial return on investment (if commercial) or societal return on investment, health or other goal (if public funder or non-commercial)**	Maximise health	Maximise financial and societal return on investment	Maximise health	Maximise financial return on investment	Maximise health
Decisions HTA designed to inform	**Broad range including:** • **Pre-clinical/preliminary market assessments** • **First estimations of pricing/reimbursement scenarios** • **Go/no go decisions** • **Technology design** • **Trial design/evidence generation strategy** • **Research prioritisation**	Go/no go decision Value proposition Technology design	Go/no go decision Value proposition Positioning of test/s Technology design Evidence required	Whether to extend testing Evidence required	Go/no go decision Value proposition Positioning of test/s Technology design Evidence required Reimbursement strategy	Whether to adopt testing Appropriate cut-off Evidence required
Decision space	**Wide including multiple:** • **Jurisdictions** • **Indications** • **Comparators** • **Funders** • **User groups** • **Thresholds (test cut-off)** • **Levels of test performance** • **Positions in pathway**	Multiple: jurisdictions levels of test performance Fixed: indication comparators position in pathway	Multiple: jurisdictions positions in pathway combinations of tests outcome measures thresholds test performance	Fixed: indication jurisdiction position in pathway test performance	Multiple: jurisdictions positions in pathway settings (primary or secondary care) combinations of tests outcome measures thresholds test performance	Multiple: test cut-off test performance Fixed: indication comparators jurisdiction

Available evidence	**Clinical studies tend to be small such that uncertainty is high** **Evidence specific to technology scarce early in the development process.** **Alternative methods of estimating parameters include:** • **Expert opinion** • **Evidence on comparators or previous generations of a technology** • **Bench or animal studies** • **Output from pharmacodynamic models** **Evidence required about usability and clinical pathways**	No evidence specific to technology.	Evidence from single retrospective study	Evidence base limited, specific to context	Evidence base limited, specific to context, based on retrospective analysis	Evidence base limited, specific to context, based on retrospective analysis
Timing	**Repeated on an iterative basis** **Pre and during development**	Pre-development. The test is hypothetical.	During development. Prototype existed and was undergoing analytical validity testing.	Molecular tests are established laboratory tests. New technology is hypothetical pathway with more extensive testing.	During development. Prototype exists – undergoing trials in a number of contexts.	Test is commercially available. New technology is hypothetical pathway incorporating test.
Business model	**Fluid -not yet defined** **Various business models available including reimbursement-based models, direct marketing to patients, clinicians or health-care organisations**	Not considered as only clinical value assessment undertaken	Multiple potential business models, may impact upon diffusion and use of technology	Extension to testing would be funded out of the National Pathology Laboratories budget	Multiple business models possible, may impact upon diffusion and use of technology	Likely fee per test – price not yet agreed.
Resources for analysis	**Often constrained at early stages due to conflicting demands on resources** **Less resource-intensive methods to establish and begin to quantify value proposition**	Constrained. 12-week project by Masters-level student and subsequent development within context of GMP Node	Constrained. Limited resource available through Glasgow Molecular Pathology Node.	Constrained. Limited resource available through Glasgow Molecular Pathology Node.	Constrained.	Constrained. No funding available.

Stance of analysis	**Positive** **Which jurisdiction, position in pathway maximises return for developers?**	Positive	Positive	Normative	Positive	Normative
Burden of proof	**Evidence credible to the development team**	Evidence credible to the development team	Evidence credible to the development team	Evidence credible to audience and the development team	Evidence needs to be credible to the development team	Evidence needs to be credible to the decision-maker

BRAF – B-Raf Proto-Oncogene, CRC – colorectal cancer, GMP – Glasgow Molecular Pathology, HTA – Health Technology Assessment, KRAS – Kirsten Rat Sarcoma, NRAS – Neuroblastoma RAS viral oncogene homolog

6.2.2 Case study 1 - Test to predict response to treatment in ovarian cancer

The empirical example which motivated this case study was a question from a clinician about whether a test to predict response to hormone treatment could be clinically/cost effective. It is interesting as it is an example of an analysis undertaken at concept stage (i.e. no research work had been undertaken) and because it is needs-driven rather than technology-driven. As ovarian cancer is an acute condition with poor prognosis, treatment is generally aggressive. A test to predict response to hormone treatment could spare some patients the burden of additional chemotherapy. The implications of a wrong test result are serious as a patient may be denied the opportunity of an effective treatment. The case study demonstrates that a simple model with limited inputs can be used to determine that investment in developing a test is not likely to be worthwhile.

6.2.2.1 Features of DF-HTA

This case study exhibits all the features of DF-HTA. The audience was an academic clinician who potentially may have started to develop the technology. The underlying objective was to maximise health. In this case the decision to be informed was whether to invest researcher resource in initiating a project and seeking funding. Technology design considerations, particularly the required sensitivity and specificity of a potential test in order to clinical value, were also of interest. Timing, resources, stance of analysis and burden of proof were all typical of DF-HTA. Business model was not relevant at this stage as the assessment only explored potential clinical value. Decision space was more limited than in many DF-HTA assessments as the development was needs-driven rather than technology-driven, so indication and position in the pathway were fixed.

6.2.2.2 Questions for consideration in DF-HTA

Table 6-3: Questions for consideration – case study 1

Feature of DF-HTA	Questions for consideration
Target audience	***Who is the analysis designed to inform?*** An academic clinician considering the research portfolio of his department at the University of Glasgow raised a number of possibilities where a test to predict treatment response may meet a clinical need. One of these possibilities was a test to predict response to hormone treatment.
Underlying user objective	***What are the developers ultimately trying to achieve through investment in development of a technology?*** ***On what basis will the developers decide whether and how it is worth continuing with the development of this technology?*** The developers were trying to meet clinical need. No research had been undertaken but if the analysis showed that a test had potential then this may have encouraged the team to seek funds and pursue this line of research. It is unclear what would constitute potential.
Decisions HTA designed to inform	***What decisions can the analysis inform?*** Whether to pursue the research project.
Decision space	***What are the possible uses of the technology?*** ***What are the most promising uses of the technology?*** ***Which of the potential use(s) should be targeted first?*** These questions are appropriate in a technology-driven assessment. In a needs-driven assessment the use is determined by the need.
Available evidence	***What evidence is available?*** ***What is the best approach to estimating parameters in the absence of evidence?*** No evidence specific to the technology is available. A rapid literature review can be used to identify appropriate estimates for the response rates to the treatments and this can be confirmed with a clinical expert. Test performance can be modelled across the full range.
Timing	**What is the most appropriate form of analysis (if any) to do now?** As patients and clinicians would be most interested in whether a test would be likely to improve response rates, a simple model to explore this question is appropriate.
Business model	**What alternative business models are possible for this technology in target jurisdictions/indications?** Not relevant to this analysis.
Resources for analysis	**What resources are available for analysis?** **What would be the most appropriate use of the resources?** No specific resources were available for this analysis so a simple model using limited evidence sources was appropriate.
Stance of analysis	**How does the analyst ensure the study meets the needs of the developers?** The analysis was presented back to a clinical expert.
Burden of proof	**Are the methods and sources of parameter estimates appropriate for this level of resources and this stage of development?** **Has the analyst communicated any limitations of the approach with the developers?** The validity of the conclusion was highly dependent on the appropriateness of the single source of response rates to hormone treatment and chemotherapy. This was communicated back to the clinical expert who was confident that the rates used were appropriate.

6.2.2.3 Does case study 1 demonstrate the features of DF-HTA?

This case study demonstrates all the features of DF-HTA. As the case study is needs-driven (i.e. the question was 'would it be worth developing a technology in this particular indication and position in the pathway?') decision space is somewhat narrower than would be envisaged in technology-driven analysis as the positioning of the technology is not being explored.

6.2.3 Case study 2 - Test to predict response to treatment in rheumatoid arthritis

The empirical example which motivated this case study was a research project at preclinical stage supported by the Glasgow Molecular Pathology Node. The technology in development was a test predicting response to two biologic treatments in rheumatoid arthritis using an algorithm. Unfortunately, the algorithm was not validated in the subsequent round of testing but the case study remains of interest from an academic perspective. Rheumatoid arthritis is a chronic condition so clinicians have more opportunity to vary treatment in order to find a therapy which works for each patient. Compared to an acute condition this may make a test less valuable as the consequences of treating with the wrong therapeutic are less severe. As the results of a small retrospective study were available, this case study provided the option to undertake analyses at two points in time to illustrate the iterative nature of DF-HTA. The case study found that continued investment was justified in the early-stage analysis but that cost-effectiveness in the mid-stage analysis was particularly sensitive to the relative price of the test, treatment and comparator. This illustrates the difficulties of developing a cost-effective test in a fast-moving therapeutic and competitive environment.

6.2.3.1 Features of DF-HTA

This case study also exhibits many features typical of DF-HTA (see Table 4-1). Again, the audience was an academic clinician and the underlying objective was to maximise health. The decision to be informed was whether to invest resource in the project and seek funding. In this case study, both clinical and economic value were important and decisions to be informed involved the positioning of the technology and the level of test performance necessary to deliver value.

The timing of this assessment was just following the retrospective analysis of a clinical trial which provided evidence of the test performance of the technology. In order to mimic the iterative approach to DF-HTA, a retrospective clinical value assessment was undertaken to demonstrate that it could be shown that the test had potential to increase response rates even before the retrospective analysis of the clinical trial was undertaken. Burden of proof and stance of analysis were both typical of DF-HTA. In this case study, the reimbursement business model was used as the method of analysis was cost-utility analysis assuming UK thresholds but sale to private payers was considered in the discussion. The decision space included multiple comparators, tests could be arranged in different orders in the clinical pathway and there were a variety of test outcomes which were possible. As the project was supported by the Glasgow Molecular Pathology Node, it would have been possible to develop a more complex health economic model including probabilistic sensitivity analysis and value of information analysis. It is, however, unclear how useful this would have been to the developers.

6.2.4 Questions for consideration in DF-HTA

Table 6-4: Questions for consideration – case study 2

Feature of DF-HTA	Questions for consideration
Target audience	***Who is the analysis designed to inform?*** A team of academics and clinicians with a technology in development supported by Glasgow Molecular Pathology Node.
Underlying user objective	***What are the developers ultimately trying to achieve through investment in development of a technology?*** ***On what basis will the developers decide whether and how it is worth continuing with the development of this technology?*** The developers were trying to meet clinical need but recognised that commercial involvement would be necessary to bring the test to market. Funding had been found to complete analytic validity testing and further funding was being sought to fund initial clinical studies.
Decisions HTA designed to inform	***What decisions can the analysis inform?*** Whether to continue to invest. Design of the technology in terms of required sensitivity and specificity. Evidence required to support value proposition such as patient and physician preferences for test outcome and likely prescribing behaviour given test results.
Decision space	***What are the possible uses of the technology?*** ***What are the most promising uses of the technology?*** ***Which of the potential use(s) should be targeted first?*** The technology comprised three tests which could be used at different points in the clinical pathway and in different chronological order in different populations. Rather than model all the alternatives the scenario with the most direct clinical relevance in the UK was assessed.
Available evidence	***What evidence is available?*** ***What is the best approach to estimating parameters in the absence of evidence?*** A retrospective analysis of a clinical trial was available to allow the performance of the test to be estimated. Guidelines and an HTA report were used to provide evidence about clinical pathways, response rates, utilities and costs to input to the model.
Timing	**What is the most appropriate form of analysis (if any) to do now?** As some clinical data was available a cost-utility analysis could be done. A retrospective analysis was also done following the simple modelling of clinical value as in case study one. The developer was also encouraged to consult stakeholders (clinicians and patients) about essential design elements of the technology (which level of response it was better to predict, whether a prediction of response to either TNFi inhibitors or rituximab or both was more useful).
Business model	**What alternative business models are possible for this technology in target jurisdictions/indications?** A cost-utility analysis was undertaken to explore potential cost-effectiveness assuming a reimbursement model with UK thresholds. Discussion included the potential of a private payer model.
Resources for analysis	**What resources are available for analysis?** **What would be the most appropriate use of the resources?** No specific resources were available, but analysis was possible as this was a Glasgow Molecular Pathology Node project. A more complex modelling approach could have been adopted (confidence intervals were available for the estimates from the small clinical trial so PSA and VOI could have been undertaken) but it is unclear whether this would have been more useful to developers at this stage of development.

Feature of DF-HTA	Questions for consideration
Stance of analysis	**How does the analyst ensure the study meets the needs of the developers?** The analysis was presented back to the developers.
Burden of proof	**Are the methods and sources of parameter estimates appropriate for this level of resources and this stage of development?** **Has the analyst communicated any limitations of the approach with the developers?** It was appropriate to use the results of the small clinical study to produce a cost-utility analysis. Developers were aware of the limitations of the approach, in particular that all the options for ordering of tests were explored. Other outcomes could not be explored as the data was not available but the bioinformatician was advised about the different analyses which would be useful and the developer advised to undertake stakeholder consultation to narrow down the preferred option.

6.2.4.1 Does case study 2 demonstrate the features of DF-HTA?

The features of this study conform to the features of DF-HTA set out in Table 4-1. Two points of interest are: the impact on decision space of whether the assessment started with the clinical need or with the technology; and the difficulty in determining whether the burden of proof is met.

This study could be seen as hybrid needs-driven and technology-driven in the sense that the initial research was based on a need to identify responders to biologic drugs in rheumatoid arthritis but as the technology was in development it was clear that there were many options to position it in the clinical pathway. This has implications for decision space. Here, the decision space is wider than study one, which was completely needs based but not as wide as the melanoma case study, where the technology in development could be used in multiple indications.

The difficulty in determining whether the burden of proof is met derives from the fact that developers may not know what they want from the DF-HTA. In this case study, the findings were reported back to the developers who concurred with the findings. Most importantly, the developers accepted the need to engage stakeholders to assist with the design and positioning of the potential test. The questions accompanying the framework may be of particular use in structuring a discussion with the developers of what they hope to achieve with the DF-HTA.

6.2.5 Case study 3 - Extension of molecular testing in colorectal cancer

In contrast to the previous two case studies, this example looked at a technology which was well established in clinical practice – testing for KRAS/NRAS and BRAF testing prior to the use of cetuximab for patients with metastatic colorectal cancer. The clinicians' question concerned whether it would be cost-effective to extend the use of the existing companion diagnostic to a broader population earlier in the diagnostic pathway. The case study is interesting as an illustration of the importance of the positioning of a diagnostic test in the existing clinical pathway to its cost-effectiveness. It demonstrated that, as a result of the downstream consequences of delayed molecular testing, the extension of molecular testing to all patients diagnosed with colorectal cancer would be likely to be cost-effective in this local context. The use of an expert steering group to provide inputs to the modelling, advise on the existing and the adapted clinical pathway and validate the models is also of interest as a pragmatic approach for local evaluations where evidence and resources may be limited.

6.2.5.1 Features of DF-HTA

As the technology is not in development it would be expected that this case study would not exhibit features typical of DF-HTA (see Table 4-1). Most features differed from those typical of DF-HTA. The audience was the decision-making body for molecular testing within NHS Scotland, the Molecular Pathology Evaluation Panel (MPEP). The underlying objective was to maximise health. The decisions to be informed were whether the extension to molecular testing was likely to be cost effective in the local context so whether the clinical team should continue to pursue the change in pathway and if they did what additional evidence may be required. The decision space was limited as in an adoption decision, with indication, position in pathway and test performance known. The clinicians sought to change the position in the pathway of an existing technology, so the timing was partway through the lifecycle of the technology. The stance of analysis was normative in that it was the cost-effectiveness thresholds set by the UK NHS which were the target. The reimbursement business model was used as the method of analysis was cost-utility analysis

assuming UK thresholds. In a number of respects, however, the features of this case study did resemble DF-HTA. The burden of proof was not equivalent to a NICE process as this decision was being made by MPEP and they have no specific methodological guidance for health economic analysis. Evidence was limited so simple modelling was undertaken with parameter estimates informed by an expert panel. Finally, resources were constrained as the project had no specific funding. It was supported through the Glasgow Molecular Pathology Node so that more complex modelling could have been undertaken but as with the previous case study, it is unclear whether a more complex approach would have provided any more value for either the clinicians looking to redevelop the pathway or MPEP as decision-makers.

6.2.5.2 Questions for consideration in DF-HTA

Table 6-5: Questions for consideration – case study 3

Feature of DF-HTA	Questions for consideration
Target audience	***Who is the analysis designed to inform?*** The clinicians developing the pathway and the Molecular Pathology Evaluation Panel.
Underlying user objective	***What are the developers ultimately trying to achieve through investment in development of a technology?*** ***On what basis will the developers decide whether and how it is worth continuing with the development of this technology?*** The clinicians developing the pathway were trying to achieve an improvement in health outcomes and patient/clinician satisfaction with the pathway. The ultimate decision-makers on the extension of testing also seek to maximise health but the testing budget is constrained and the impact on cost of the change in pathway is an important factor in their decision-making.
Decisions HTA designed to inform	***What decisions can the analysis inform?*** The demonstration of the economic impact of the proposed change in the testing pathway informed the clinicians' decisions about whether to continue to push for the change and what additional evidence may be necessary to evidence the impact.
Decision space	***What are the possible uses of the technology?*** ***What are the most promising uses of the technology?*** ***Which of the potential use(s) should be targeted first?*** The decision space in this case study was narrow as the indication and position in pathway for this technology was constrained by the aim of the study.
Available evidence	***What evidence is available?*** ***What is the best approach to estimating parameters in the absence of evidence?*** Local evidence was required. Evidence was taken from national sources and local clinical audit. Where required estimates for parameters were provided by a multi-disciplinary expert group and varied widely in sensitivity analysis.

Feature of DF-HTA	Questions for consideration
Timing	**What is the most appropriate form of analysis (if any) to do now?** The most appropriate form of analysis was a cost-minimisation analysis. It was important for the ultimate decision makers that this extension in testing should save costs overall and there was no likelihood of health outcomes being adversely affected by testing being extended to all patients with colorectal cancer rather than only those with metastatic disease.
Business model	**What alternative business models are possible for this technology in target jurisdictions/indications?** A reimbursement business model was appropriate as the setting was the NHS in Scotland.
Resources for analysis	**What resources are available for analysis?** **What would be the most appropriate use of the resources?** No specific resources were available, but analysis was possible as this was a Glasgow Molecular Pathology Node project. A more complex modelling approach could have been adopted (but uncertainty would have had to be estimated through expert elicitation for many of the parameters). It is unclear whether a more complex modelling approach would have been more useful to developers.
Stance of analysis	**How does the analyst ensure the study meets the needs of the developers?** The analysis was presented back to the developers and to MPEP, the ultimate decision makers. The approach to modelling and evidence sources used appeared sufficient for the decision-maker. However, circumstances around molecular testing for colorectal cancer changed as micro satellite instability testing was required earlier for some diagnosed patients so additional modelling would have been required before a definitive decision could be made.
Burden of proof	**Are the methods and sources of parameter estimates appropriate for this level of resources and this stage of development?** **Has the analyst communicated any limitations of the approach with the developers?** The approach to modelling and evidence sources used appeared sufficient for the decision-maker. However, circumstances around molecular testing for colorectal cancer changed as micro satellite instability testing was required for all diagnosed patients so additional modelling would have been required before a definitive decision could be made.

6.2.5.3 Does case study 3 demonstrate the features of DF-HTA?

This case study does not exhibit most of the features of DF-HTA identified in Table 4-1. The ultimate audience for the assessment was a public payer and the main decision to be informed was whether to extend the testing given the estimated economic impact. The decision space was narrow as the indication, jurisdiction and position in pathway were all fixed. The timing of the assessment was when the technology was well established in clinical use. The stance of analysis was normative, as the technology was assessed according to the decision-maker's criteria. This contrasts with the exploratory positive stance taken in DF-HTA when the analysis seeks the most beneficial positioning and design for the technology.

Initially, the justification for including this case study and the Prostate Health Index case study was that the clinicians were effectively developing a new pathway including an established technology. Whilst this is undoubtably true, it is clear that this kind of HTA is different in most features from what has been termed DF-HTA. The similarities between DF-HTA and this case study are that resources were somewhat limited and that evidence was lacking and needed to be supplemented with input from an expert panel. These similarities were driven by the context-specific nature of the assessment, as although evidence was available on a national basis, it was the local pathway information which needed to be supplemented with clinical audit or expert input. This difficulty in establishing relevant clinical pathways is discussed by Abel et al (2019) in connection with diagnostic technologies assessed by the DECs. Resource limitation is also typical of local assessment as dedicated support is not available as it may be in a national assessment programme.

Burden of proof is, again, a difficult feature to judge for this assessment. MPEP have no reference case and applications for funding are generally not supported by a formal health economic analysis. It is unclear what evidence the panel base their decisions on. The panel accepted the evidence as presented but it did not lead immediately to a change of practice as an additional test was being assessed which impacted upon this clinical pathway. A full scoping of the decision problem may have highlighted this additional test at an earlier stage, and it could have been incorporated into the analysis.

6.2.6 Case study 4 - Diagnostic test in malignant melanoma

Rather than looking at tests to predict treatment response as in the previous three case studies, this case study looks at an imaging test for the diagnosis of malignant melanoma. There was a prototype of the technology in existence and the developers had completed one clinical study but had not finalised placement in the clinical pathway. This case study concerns an acute condition which is the subject of government targets in that all patients with suspected skin cancer should be seen within two weeks. This creates capacity constraints in the clinical pathway. The case study found that costs could be saved by introducing a test to triage between referral for a suspicious mole and specialise dermatologist review. These cost savings were limited as only the

cost of a consultation was saved. The test needed to be highly sensitive as it was ruling out melanoma (Baeyens et al, 2019). Of more interest, if capacity constraint of specialist dermatologists was built into the model, the triage test is able to 'enrich' the population (i.e. increase the prevalence of malignant melanoma in the population referred to the specialist dermatologist) so that more cases are diagnosed overall.

6.2.6.1 Features of DF-HTA

This case study exhibits many features typical of DF-HTA (see Table 4-1). The audience is an SME commercial developer with this single technology in development. The underlying objective was to maximise financial return on investment. The decisions to be informed were whether it was worthwhile to continue to invest resource in the project given an optimal position in the clinical pathway, how the technology should be designed to maximise potential in this position, the evidence required to demonstrate value and which business model had the most potential. The timing of this assessment was just following a small clinical trial which provided evidence of the test performance of the technology. Burden of proof and stance of analysis were both typical of DF-HTA. In this case study, the threshold from the UK reimbursement business model was used as headroom calculations were undertaken. Sale to private payers was considered in the discussion. The decision space included multiple comparators depending on the position in the clinical pathway and the setting. The developers were prepared to provide more resource for analysis, so it would have been possible to develop a more complex health economic model including probabilistic sensitivity analysis and value of information analysis. However, given the uncertainty around positioning of the technology in the clinical pathway, it is unclear how useful this would have been to the developers.

6.2.6.2 Questions for consideration in DF-HTA

Table 6-6: Questions for consideration – case study 4

Feature of DF-HTA	Questions for consideration
Target audience	***Who is the analysis designed to inform?*** An SME commercial developer and investors.
Underlying user objective	***What are the developers ultimately trying to achieve through investment in development of a technology?*** ***On what basis will the developers decide whether and how it is worth continuing with the development of this technology?*** The developers were trying to bring the product to market and maximise financial return on investment. Private investment had funded development and clinical testing to date, but further funding was required for ongoing development and further clinical testing.
Decisions HTA designed to inform	***What decisions can the analysis inform?*** Whether to continue to invest. Value proposition for the technology. Positioning of technology. Design of the technology in terms of required sensitivity and specificity. Evidence required to support value proposition.
Decision space	***What are the possible uses of the technology?*** ***What are the most promising uses of the technology?*** ***Which of the potential use(s) should be targeted first?*** The technology could be used in different settings and the cut-off levels could be set to detect different levels of skin abnormality. At its most restrictive it would detect only malignant melanoma but there was also scope to detect other skin cancers and abnormalities.
Available evidence	***What evidence is available?*** ***What is the best approach to estimating parameters in the absence of evidence?*** A small clinical trial was available to provide estimates of test performance. Guidelines and other literature sources provided information about clinical pathways. Publicly available data was used to estimate costs.
Timing	**What is the most appropriate form of analysis (if any) to do now?** As positioning was uncertain, it was not appropriate to develop a detailed model. Rather, a simple model which could demonstrate the costs and consequences of using the technology in different settings was developed. The developer was also encouraged to consult and potentially collaborate with a wide range of stakeholders (clinicians and commissioners) about essential design elements of the technology (which settings would be the most appropriate and what functionality would be required in those settings).
Business model	**What alternative business models are possible for this technology in target jurisdictions/indications?** A cost-consequence model assumed a UK setting. Discussion included the potential of other models.
Resources for analysis	**What resources are available for analysis?** **What would be the most appropriate use of the resources?** No resource was provided but developers would have paid for a more extensive analysis. The value of more complex health economic modelling was unclear given the stage of development. More extensive stakeholder consultation and collaboration would likely have been a better investment of constrained development resources.

Feature of DF-HTA	Questions for consideration
Stance of analysis	**How does the analyst ensure the study meets the needs of the developers?** The analysis was presented back to the developers. They did not feel the analysis met their needs. They were hoping for a more detailed analysis of the relevant epidemiology. However, the granular level of epidemiological data which they required is not publicly available. Hence, the recommendation to collaborate closely with a range of healthcare providers in developing context-specific value propositions.
Burden of proof	**Are the methods and sources of parameter estimates appropriate for this level of resources and this stage of development?** **Has the analyst communicated any limitations of the approach with the developers?** It was appropriate to undertake a simple modelling study to link the costs avoided later in the clinical pathway to the cost and performance of the test. A significant component of the value proposition for the technology in the UK is that it protects some specialist dermatologist capacity, which is constrained. This aspect was built into the modelling to illustrate that even a test with relatively low performance can improve case detection by enriching the prevalence in the population accessing the specialist dermatologist.

6.2.6.3 Does case study 4 demonstrate the features of DF-HTA?

The features of this study align with all the features of DF-HTA identified in Table 4-1. As the assessment was technology-driven positioning of the technology was important and decision space was wide. Stance of analysis is of interest as it is difficult to determine whether this assessment met the needs of the developers. The case studies were completed before the questions set out in Table 4-2 were drafted. On completion of the assessment it became clear that the analysis did not meet the developers' expectations even if it may have met their needs. The developers' expectation was that the study would produce a full cost-effectiveness analysis using detailed data on incidence and clinical pathways. However, at the current stage of development, positioning was still sufficiently uncertain that such a model would have been of questionable value. At this stage, more focus on developing value propositions in collaboration with a range of stakeholders in a variety of settings was of more value to the developers. There is evidence in the literature of developers' expectations or actions not being aligned with analysts' conclusions (Miquel-Cases, 2016a, Markiewicz, 2017b) thus highlighting the need for all parties to be clear about the purpose of the assessment and the appropriateness of different approaches at different points in the analysis. Use of the questions in Table 4-2 may have made this misunderstanding explicit.

6.2.7 Case study 5 - Diagnostic test in prostate cancer

The final case study looks at a second diagnostic test, this time in prostate cancer. The technology was on the market and a number of clinical studies had been completed. Again, this case study concerned an acute condition where the consequences of missing clinically significant cases could be severe. The current diagnostic pathway had room for improvement as the Prostate Specific Antigen (PSA) blood test identified a large number of false positives so patients then underwent prostate biopsy, an invasive and expensive procedure. The study quantified the cost savings and consequences ensuing from the introduction of Prostate Health Index (PHI) as a triage test after PSA before biopsy at different thresholds of sensitivity and specificity. The case study showed that biopsy and its attendant complications are sufficiently costly that the introduction of a triage test which avoids some biopsies has the potential to be cost effective. A high sensitivity test is required to rule out clinically significant disease, but it is specificity which drives cost savings through avoided biopsies.

6.2.7.1 Features of DF-HTA

As the technology was available on the commercial market, it would be expected that this case study would not exhibit features typical of DF-HTA (see Table 4-1). Most features differed from those typical of DF-HTA. The audience was the clinicians developing the potential care pathway as well as the ultimate decision-making body for Hong Kong. The underlying objective was to maximise health. The decision to be informed was whether the inclusion of PHI in the diagnostic pathway was likely to be cost effective in the local context. The decision space was limited as in an adoption decision, with indication and position in pathway known. The study explored the optimum thresholds for sensitivity and specificity to adopt using cost-consequence analysis. The stance of analysis was normative in the sense that the criteria for adoption would be externally determined by the decision-making authority. The reimbursement business model was assumed with costs and care pathways appropriate to the Hong Kong healthcare authority. In a number of respects the features of this case study did resemble DF-HTA. In terms of burden of proof, no specific methodological guidance for health economic analysis was available or followed. Evidence was limited to a single clinical study and deterministic modelling was

undertaken. Finally, resources were constrained as the project had no specific funding.

6.2.7.2 Questions for consideration in DF-HTA

Table 6-7: Questions for consideration – case study 5

Feature of DF-HTA	Questions for consideration
Target audience	***Who is the analysis designed to inform?*** The clinicians developing the pathway and Hong Kong Hospital Authority.
Underlying user objective	***What are the developers ultimately trying to achieve through investment in development of a technology?*** ***On what basis will the developers decide whether and how it is worth continuing with the development of this technology?*** The clinicians developing the pathway and the ultimate decision makers were trying to achieve an improvement in health outcomes and reduction in costs. If similar health benefits could be achieved with the potential for reduced costs it would be worth continuing with the development of the technology.
Decisions HTA designed to inform	***What decisions can the analysis inform?*** The demonstration of the economic impact of the proposed change in the testing pathway informed the clinicians' decisions about whether to continue to push for the change and what additional evidence may be necessary to evidence the impact.
Decision space	***What are the possible uses of the technology?*** ***What are the most promising uses of the technology?*** ***Which of the potential use(s) should be targeted first?*** The decision space in this case study was narrow as the indication and position in pathway for this technology was constrained by the aim of the study. However, there was still flexibility over the choice of test performance thresholds and this was explored in the analysis.
Available evidence	***What evidence is available?*** ***What is the best approach to estimating parameters in the absence of evidence?*** Local evidence was required. Evidence was taken from a single clinical effectiveness trial and local cost data.
Timing	**What is the most appropriate form of analysis (if any) to do now?** The most appropriate form of analysis was a cost-consequence analysis. It was important for the ultimate decision makers that costs savings should be made. It was important to assess the level of cost savings available in light of the potential consequences, particularly in terms of missed cases.
Business model	**What alternative business models are possible for this technology in target jurisdictions/indications?** A reimbursement business model was appropriate as the setting was the Hong Kong Hospital Authority.
Resources for analysis	**What resources are available for analysis?** **What would be the most appropriate use of the resources?** No specific resources were available. The analysis undertaken was relatively resource light and was a useful step in demonstrating to the clinicians and the ultimate decision maker the level of cost savings available and the potential consequences of different threshold choices.
Stance of analysis	**How does the analyst ensure the study meets the needs of the developers?** The analysis was presented back to the developers. The developers continued the research effort and have since produced a full cost-effectiveness model comparing PSA and PHI as an initial test in opportunistic screening of men in primary care (Teoh et al, 2020).

Feature of DF-HTA	Questions for consideration
Burden of proof	**Are the methods and sources of parameter estimates appropriate for this level of resources and this stage of development?** **Has the analyst communicated any limitations of the approach with the developers?** The approach to modelling and evidence sources used appeared sufficient for the decision-maker.

6.2.7.3 Does case study 5 demonstrate the features of DF-HTA?

As with the assessment of the extension to molecular testing in colorectal cancer (case study 3) this study differs from DF-HTA in most of the features. The ultimate audience was a public payer and the main decision to be informed was whether to adopt a commercially available test. As in case study 3, the features which were consistent with DF-HTA were evidence limitations and resource constraint. Again, both these features were driven in this case study by the local nature of the assessment. Evidence at the granular level and care pathways were very specific to this setting and were derived from a single clinical study. Resource for the study was not provided through an evaluation body. For burden of proof, it was not clear what level of evidence decision-makers required to make a decision. The analysis was useful to demonstrate the likely cost-effectiveness of the introduction of PHI and may have encouraged further research into the cost-effectiveness of the test in other positions in the pathway by the Hong Kong based clinical team (Teoh et al, 2020).

6.2.8 Do the case studies demonstrate the features set out in the framework?

Three of the case studies exhibit all the features of DF-HTA as identified in Table 4-1. The remaining two case studies exhibit few of the features (including limited evidence and restrained resources) but these case studies are quite different as they assess existing technologies for public payer audiences. The framework of features does seem to describe DF-HTA adequately. The questions presented in Table 4-2 appear to identify useful aspects of the assessment and may have been helpful to make explicit features of the analysis at the outset of each of the case studies. In particular, questions about the stance of analysis and burden of proof may help to make explicit the nature of the assessment and

ensure that both the developer and the analyst were clear about the goals of the analysis.

6.3 Activities and methods used in the case studies

The generic process of development-focused HTA presented in Figure 5-1 includes the core activities of clinical and economic value assessment. These assessments take place in an iterative process and each activity is likely to be revisited on more than one occasion. Table 6-8 sets out the activities and methods of DF-HTA selected in the case studies. The first step in all case studies was to undertake a clinical value assessment using a simple modelling approach to map clinical pathways with and without the new technology. In cases where the clinical value assessment was positive the next step was to undertake an economic value assessment, again using simple decision tree models to undertake cost minimisation, cost consequence or cost utility analysis to estimate headroom.

Table 6-8: Activities and methods of DF-HTA used in case studies

	Case study 1 Test of response to treatment in ovarian cancer	**Case study 2 Test of response to treatment in rheumatoid arthritis**	**Case study 3 Extension of molecular testing in colorectal cancer**	**Case study 4 Diagnostic test in malignant melanoma**	**Case study 5 Diagnostic test in prostate cancer**
Activities of development-focused HTA	Clinical value assessment	Clinical and economic value assessments	Clinical and economic value assessments	Clinical and economic value assessments	Clinical and economic value assessments
Methods used	Model of clinical pathways using decision tree	Model of clinical pathways using decision tree Cost-utility analysis	Model of clinical pathways using decision tree Cost-minimisation analysis	Model of clinical pathways using decision tree Cost-consequence analysis	Model of clinical pathways using decision tree Cost-consequence analysis

DF-HTA – development-focused HTA, HTA – health technology assessment

7 Case study 1 - Test and Treat Superiority Plot: a simple tool for developers of tests for treatment response.

7.1.1 Introduction

Tests, such as companion diagnostics, that attempt to predict an individual's response to a specific treatment are likely to be an important component of precision medicine. For developers of medical technologies, it is useful to articulate a value proposition at an early stage of development (Rogowski et al, 2016) and at this stage developers often do not have the time or funds to fully evaluate the potential impact of the test in development (Vallejo-Torres et al, 2008). However, patients, physicians, and by extension, developers of diagnostic technologies are most likely to be interested in those tests that improve clinical outcomes, so a decision model that predicts net clinical benefit may be sufficient for decision-making during early development. The objectives of this chapter are to develop a generic decision analytic model, predicting probability of response, comparing 'test and treat' and 'treat all' strategies for a test that predicts individual response to a given treatment and to introduce the Test and Treat Superiority Plot which visually presents the combinations of sensitivity and specificity at which a test and treat strategy would improve expected treatment response when compared with a treat all strategy.

The generic approach is presented then applied in a case study of response to hormone treatment or chemotherapy in ovarian cancer.

This chapter is based upon an MSc dissertation submitted Dmitry Ponomarev and co-supervised by Janet Bouttell and Neil Hawkins. The chapter forms the basis of an article (using a different case study) in preparation.

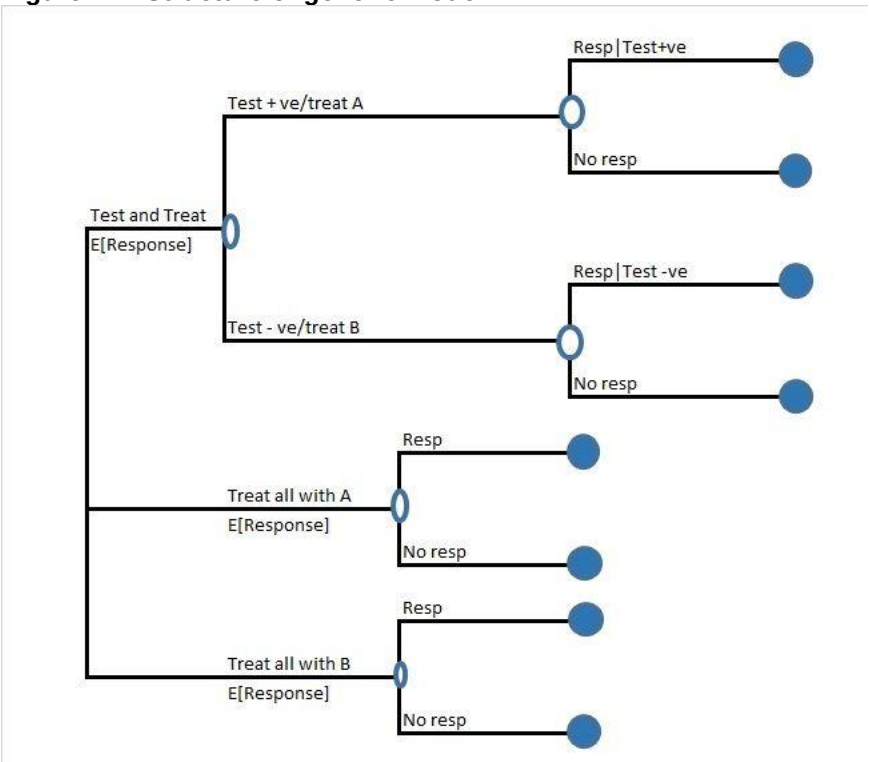
7.1.2 Method

A decision-tree model was developed comparing the two 'treat all' strategies with either treatment A or treatment B with a test and treat strategy where patients are tested for response to treatment A and are treated with A if the test is positive (and B if test negative). It is assumed that the test is not

predictive of response to treatment B. Extensions to this model are considered in the discussion.

The model is shown in Figure 7-1. The first strategy is the test and treat strategy where all patients are tested. The patients with a positive test result (which indicates that they are likely to be responders to treatment A) are treated with treatment A. A proportion of patients treated with treatment A respond (determined by the positive predictive value of the test and the response rate to treatment A - pA) and the remainder do not respond. Those that are predicted not to respond are treated with Treatment B with response rate pB (this response rate is assumed to be independent of the test result). The other two branches of the decision tree show treat all strategies using either treatment A with response rate pA or treatment B with response rate pB. Response rates to the two treatments are known.

Figure 7-1: Structure of generic model



+ve – test positive (i.e. indicates likely responder to treatment A), Resp – responder/response, -ve – test negative (i.e. indicates likely non-responder to treatment A, E – estimate

Based on this decision tree it can be shown that, for a given specificity, the threshold sensitivity at which the expected response of the 'test and treat' strategy for a test predicting individual response to treatment A equals the optimal 'treat all' strategy can be estimated based solely on the odds ratio of

response comparing treatment A and B and is independent of the absolute probability of response. The equations estimating the threshold sensitivity are given below.

If the Odds Ratio for treatment A compared with treatment B is >= 1:

$$\text{sensitivity} = 1 - \text{specificity}/\text{OR}_{\text{a.vs.b}} \qquad \text{eqn. 1}$$

If the Odds Ratio for treatment A compared with treatment B is <= 1:

$$\text{sensitivity} = (1 - \text{specificity})/\text{OR}_{\text{a.vs.b}} \qquad \text{eqn. 2}$$

The mathematical proof is provided as Appendix 10.

The threshold sensitivity is plotted as a function of specificity for a given odds ratio. The plot has been termed a 'Test and Treat Superiority' Plot. Specificity is plotted on the X axis in reverse (1 to 0) so that the plot is analogous to a Receiver Operator Characteristics (ROC) curve. It is also possible to chart the expected net clinical gain with the test and treat strategy compared with the optimal treatment strategy. This is a function of the expected responses to treatment A and treatment B. The plot and chart of clinical gain are presented in the results to the case study.

RShiny is a software package which allows users to use the statistical software R to build interactive web-based apps. An RShiny app has been developed, based on the plot and table in this chapter, and it is available at https://nshpublicapps.shinyapps.io/test_and_treat_app/.

7.1.3 Case study

7.1.3.1 Background

Ovarian cancer (OC) is the fourth most common cause of death due to cancer in women (Denny, 2013) with the majority of cases presenting at an advanced stage (SEER, 2014). Epithelial OC, where the cancer starts from the cells covering the outer surface of the ovary, accounts for up to 90% of all OC (Granstrom et al, 2008; McCluggage, 2008). Prognosis is generally poor, with

average 5-year survival below 40% (de Angelis et al, 2017; Usach et al, 2015) and relapse rates within three years of treatment around 70% (Ledermann et al, 2013). OC is generally treated aggressively following diagnosis. For Stage 1, OC National Institute for Health and Care Excellence (NICE) guidelines in the UK include surgical staging with possible subsequent chemotherapy depending on the results of staging and other factors (NICE, 2011). For advanced (stage II-IV) cancer primary debulking surgery, followed by chemotherapy (adjuvant therapy) is recommended (NICE, 2011). Additionally, chemotherapy could precede surgery (neoadjuvant chemotherapy). Standard chemotherapy regimens include platinum-based drugs (cisplatin or carboplatin) and/or cytotoxic drugs, such as paclitaxel, as first-line therapy.

Response to treatment in OC is most commonly assessed using the response evaluation criteria in solid tumours (RECIST) version 1.1 (Eisenhauer et al, 2008). This guidance distinguishes complete response, partial response, stable disease or disease progression. Either complete response or partial response may be considered best overall response (Eisenhauer et al, 2008). The progression-free period after first-line treatment is considered to be one of the most important factors associated with probability of response to second-line therapy and subsequent rounds of treatment (Ledermann et al, 2013). Patients are categorised as “platinum-refractory” if progression occurs during treatment or within four weeks of the last dose, “platinum-resistant” if progression occurs within six months or “platinum-sensitive” if progression occurs after more than six months. Platinum-resistant and refractory patients generally receive palliative chemotherapy as their survival is unlikely to exceed 12 months (Ledermann et al, 2013). Platinum-sensitive patients receive additional lines of platinum-based therapy in combination with cytotoxic drugs but eventually become resistant to any chemotherapeutic regimen (Ledermann et al, 2013). This drives the search for new agents and treatment strategies. Hormone therapy (HT) is considered one such treatment strategy (Paleari et al, 2017).

Hormone treatment is not routinely used to treat patients with refractory OC. Plausible biological mechanisms exist whereby gonadal steroids may impact upon OC and to date, a variety of hormonal agents have been used in clinical studies, (including anti-estrogen tamoxifen, aromatase inhibitors, anti-androgens,

progestins and luteinising hormone releasing agonists). A recent comprehensive review and meta-analysis of clinical benefit and risk of death with endocrine therapy included 53 trials and 2,490 patients with OC (Paleari et al, 2017). Populations were a combination of platinum-resistant, not platinum resistant, a mixture of both or platinum-resistance was not reported. The summary clinical benefit rate (SCBR, defined as the proportion of patients who achieved complete response, partial response or stable disease or the proportion of patients with no disease progression within the study period) for any HT was 0.41 (95% confidence interval (CI), 0.34-0.48) (Paleari et al, 2017). However, the authors acknowledged a high level of heterogeneity of studies within the analysis. As well as the mix of platinum-resistance many treatments and comparators were included. Only one single randomized control trial was identified which directly compared the two forms of treatment considered in the case study (Lindemann et al, 2017). This was preferred as the source of the base case model estimates as it was a relatively recent trial which was specific to the population, technology and comparators in the case study. A clinical expert confirmed that this trial was representative of expected response rates in this population.

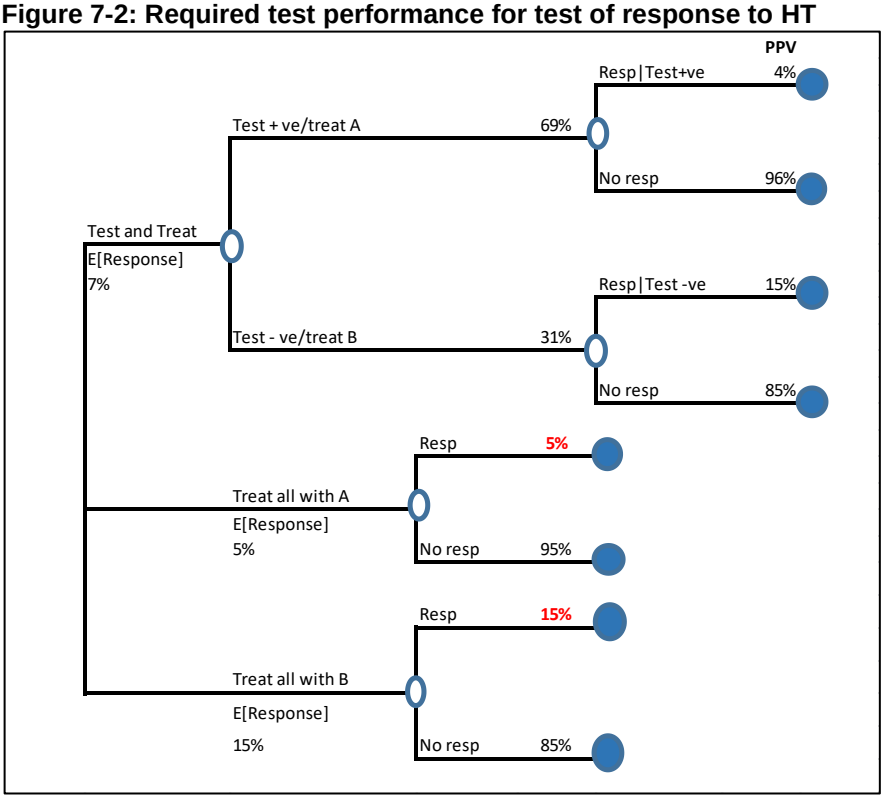
Given the high level of heterogeneity in the population of patients with OC, it is possible that there is a sub-population of patients who are likely to respond well to HT. If a biomarker test could be developed to identify such patients this could improve overall response rates and overall survival. The aims of this case study were to explore the test performance necessary in order for a test and treat strategy to deliver improved clinical performance over strategies to treat all with either chemotherapy or hormone therapy in a population of platinum-refractory/resistant patients with OC. This information would also form the basis of a go/no go decision, if the hypothetical test was unable to demonstrate sufficient clinical value to justify the costs of investment even with perfect test performance.

7.1.3.2 Method

The case study model is shown in Figure 7-2. The first strategy is the test and treat strategy where all patients are tested for response to hormone treatment. Patients with a positive test result are treated with hormone therapy. A proportion of patients treated with hormone therapy respond (determined by

the positive predictive value of the test and the response rate to hormone therapy - pA) and the remainder do not respond. Those that are predicted not to respond are treated with chemotherapy with response rate pB. This response rate is assumed to be independent of the test result. The other two branches of the decision tree show treat all strategies using either hormone therapy with response rate pA or chemotherapy with response rate pB.

Figure 7-2: Required test performance for test of response to HT



+ve – test positive (i.e. indicates likely responder to treatment A), HT – hormone treatment, Resp – responder/response, -ve – test negative (i.e. indicates likely non-responder to treatment A, E – estimate, PPV – positive predictive value

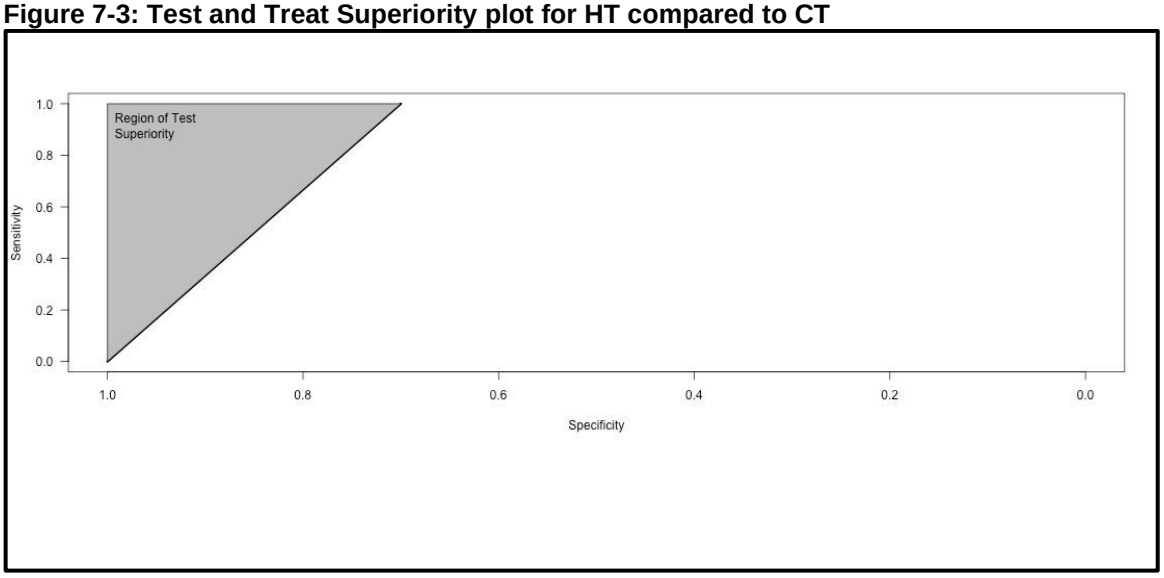
A single randomized control trial was used as the source of the model inputs as it was a recent trial directly comparing the two forms of treatment which are the subject of the case study (Lindemann et al, 2017). It estimated response to hormone therapy in platinum-resistant patients (treatment A) at 5% and response to chemotherapy (treatment B) at 15%. From these values an odds ratio can be calculated for treatment A compared to B of 0.31.

A range of odds ratios (with both chemotherapy and hormone therapy having better response rates) were explored to illustrate the impact of this on the required test performance.

7.1.3.3 Results

For a test and treat strategy to improve response rates in the base case scenario (response to CT - 15% and odds ratio 0.31) specificity must be above 69% at 100% sensitivity. As specificity increases required sensitivity decreases until it equals 0% at 100% specificity (see Figure 7-3).

Figure 7-3: Test and Treat Superiority plot for HT compared to CT



CT – chemotherapy, HT – hormone treatment

As the response rate to HT is not high in the base case (5%) even a perfect test with 100% sensitivity and specificity would only deliver a 4% improvement in overall response rates (see Figure 7-4). As the odds ratio increases beyond one (i.e. hormone treatment has higher response rate relative to chemotherapy) the proportion of responders overall increases but the benefit of the test and treat strategy is compared to a higher response rate so does not increase beyond that achieved when the treatments are equally effective in terms of response rates.

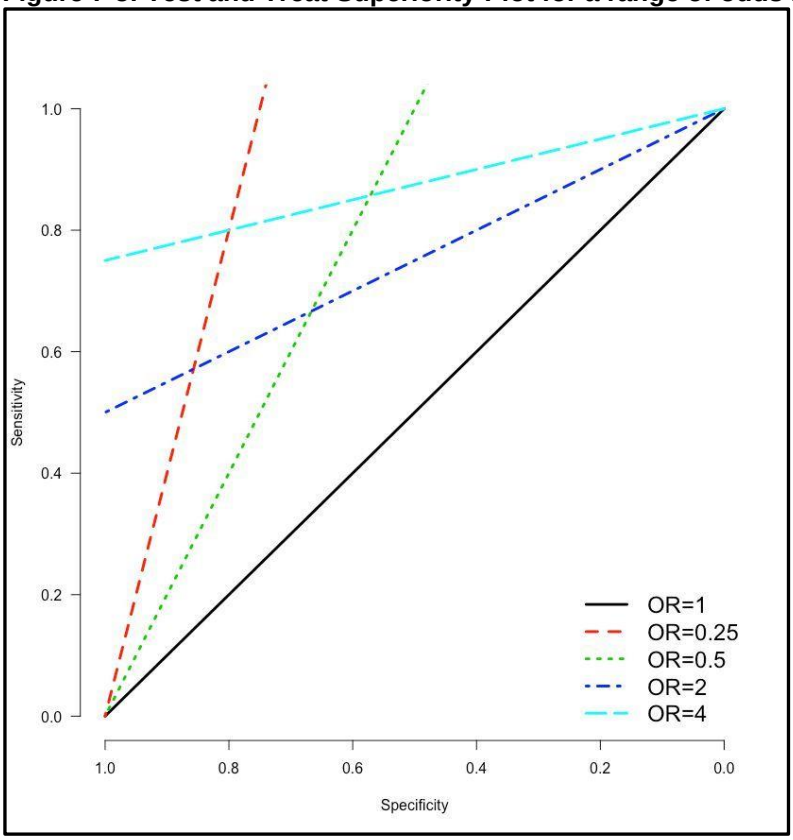
Figure 7-4: Gain from test and treat strategy

	Spec= 1	Spec= 0.9	Spec= 0.8	Spec= 0.7	Spec= 0.6	Spec= 0.5	Spec= 0.4	Spec= 0.3	Spec= 0.2	Spec= 0.1
Sens= 1	4%	3%	1%	0%	0%	0%	0%	0%	0%	0
Sens= 0.9	4%	2%	1%	0%	0%	0%	0%	0%	0%	0
Sens= 0.8	3%	2%	1%	0%	0%	0%	0%	0%	0%	0
Sens= 0.7	3%	2%	0%	0%	0%	0%	0%	0%	0%	0
Sens= 0.6	3%	1%	0%	0%	0%	0%	0%	0%	0%	0
Sens= 0.5	2%	1%	0%	0%	0%	0%	0%	0%	0%	0
Sens= 0.4	2%	0%	0%	0%	0%	0%	0%	0%	0%	0
Sens= 0.3	1%	0%	0%	0%	0%	0%	0%	0%	0%	0
Sens= 0.2	1%	0%	0%	0%	0%	0%	0%	0%	0%	0
Sens= 0.1	0%	0%	0%	0%	0%	0%	0%	0%	0%	0

Sens – sensitivity, Spec - specificity

Figure 7-5 illustrates the threshold sensitivity and specificity required for a test and treat strategy to dominate the optimum treat all strategy for a range of odds ratios. The area above and to the left of the line is the area where test and treat strategy improves overall response rate.

Figure 7-5: Test and Treat Superiority Plot for a range of odds ratios



OR – odds ratio. Area above and to the left of the line is the area where test and treat strategy improves overall response rate.

7.1.3.4 Discussion

This chapter has shown that based on only the odds ratio comparing two treatments, it is possible to identify the threshold at which a test and treat strategy for a test (predicting response for one of treatment) outperforms a treat all strategy using the optimum treatment. This threshold varies only with the odds ratio between the treatments rather than with the level of treatment response per se. The Test and Treat Superiority Plot illustrates the threshold above which a test and treat strategy will increase overall response rates and thus indicates the minimum levels that a test must achieve in order to begin to establish a value proposition for a potential diagnostic technology.

Table 7-1 shows how sensitivity and specificity can be estimated for the form of test under consideration. The proportion of patients meeting response criteria could be taken from empirical data and the test result determined retrospectively from data samples.

Table 7-1: Estimation of sensitivity and specificity

Prediction of Response to Hormone Treatment	Test +ve	Test -ve
Patient meets response criteria post treatment A	0.25	0.05
Patient does not meet response criteria post treatment A	0.25	0.45
Sensitivity (0.25/0.3)		0.83
Specificity (0.45/0.7)		0.64

-ve – negative, +ve – positive

A strength of this approach is that it is simple and easy to operationalise. It is clear to developers the levels of test performance they must achieve in order for a testing strategy to be preferred to a treat all strategy. The Test and Treat Superiority Plot is novel although it is similar to test-threshold graphs presented in Hunink et al (2014). Hunink et al (2014) plotted the pre-test and post-test probabilities of disease for patients with positive and negative test results in order to derive a threshold for the pre-test probability of disease below which you would neither test or treat and a further threshold above which you would treat regardless of a test result. Hunink et al's plot involves test performance, disease prevalence and a previously derived treatment threshold (where an individual would be indifferent between the treat or no treat strategies) so is more complex that the Test and Treat Superiority plot.

A key limitation of this simple model is the assumption that expected probability of response to the second treatment is independent of the test result. The decision model could easily be extended to tests where the expected response to both treatments varies conditional on test results. It should be noted that the evidence requirements are more extensive in this situation.

However, as the approach is intended as a first step in assessing potential clinical value the simple model may still be of value as it may allow the development of technologies that are unlikely to add value even with strong assumptions to be terminated at an early stage. Although some forms of health technology assessment require a systematic approach to the identification of relevant literature and potentially meta-analysis to synthesise the evidence, this very early form of assessment of potential value requires less onerous forms of evidence collection. Expert opinion could be used as could a very rapid review of the literature to identify an appropriate starting point for the analysis, as exemplified here. Care must be taken to consider whether there is any limitation in the availability of the treat all strategies. For example, many expensive cancer treatments are not available on a treat all basis. In this case, the model may still be used but the comparison of the test and treat response rate should be against the treat all comparator which is available.

Another limitation is the focus on response rates rather than quality of response, but again the same logic applies that this technique can be a useful first step in evaluation for developers or funders of health technologies rather than providing definitive guidance. This technique would not be intended to replace any further clinical or cost-effectiveness modelling rather as a very early technique to assess whether the technology is likely to have clinical utility.

This Test and Treat Superiority Plot offers a simple and quick first assessment tool for a test of response to a treatment. The data requirement is low and the model is sufficiently flexible to allow numerous scenarios to be assessed in a short space of time. The visual presentation of the plot facilitates discussion within development teams and between developers and potential funders.

8 Case study 2 - Development-focused HTA of test to predict response to biologic treatments in rheumatoid arthritis

8.1 Introduction

Rheumatoid arthritis (RA) is a chronic inflammatory disease characterised by progressive, irreversible joint damage, impaired joint function, and pain and tenderness caused by swelling of the synovial lining of joints (Stevenson et al, 2016). RA affects approximately 1% of the population (SIGN, 2011), is more prevalent in women than men and the typical age at onset is between 40 and 80 with peak incidence in the 70s (Stevenson et al, 2016). Diagnosis of RA is based on guidelines known as the Rheumatoid Arthritis Classification Criteria jointly issued by the American College of Rheumatology (ACR) and the European League Against Rheumatism (EULAR) (Stevenson et al, 2016). These guidelines focus on early signs of persistent and/or erosive disease such as joint involvement, duration of symptoms, serology (presence of certain antibodies in the serum) and acute-phase reactants (proteins which respond to the presence of inflammatory agents) (Stevenson et al, 2016) as well as the absence of an alternative diagnosis. It is important to start treatment in response to early disease due to the irreversibility of the joint damage if RA is left untreated.

In the UK, monitoring of disease-progression is often undertaken using the Disease Activity Score 28 joints (DAS28). This assesses 28 joints in terms of swelling and tenderness, incorporates a measure of erythrocyte sedimentation rate (ESR – which measures how quickly red blood cells settle at the bottom of a test tube) and a patient-reported subjective assessment of disease activity. EULAR criteria use the individual change in DAS28 and the level of DAS28 reached to classify patients as responders, good responders or non-responders (Stevenson et al, 2016). Table 8-1 sets out the change and levels of DAS28 corresponding to each response level. Another outcome measure widely used in RA is the Health Assessment Questionnaire (HAQ). This is a patient completed disability assessment comprising a score from 0-3 in increments of 0.125 where a higher score indicates higher disability. In economic evaluations HAQ has been shown to correlate to utility scores such as those gathered using EQ-5D (Euroqol

5 dimensions - which is a standardized instrument for measuring generic health status) (Stevenson et al, 2016).

Table 8-1: Determining EULAR response based on DAS28

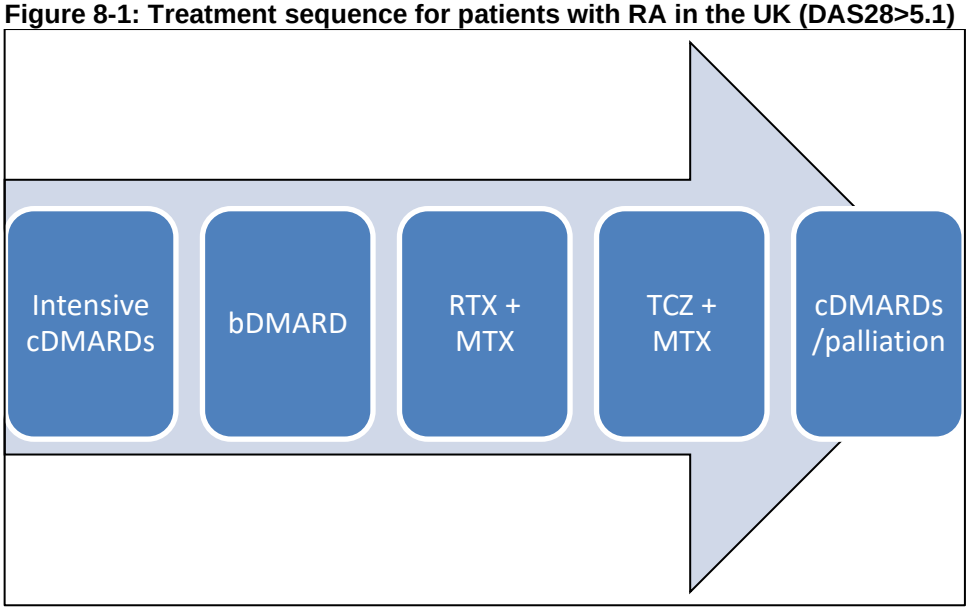
	Improvement in DAS28		
DAS28 at end point	>1.2	>0.6 and <=1.2	<= 0.6
<= 3.2	Good	Moderate	Non
>3.2 and <=5.1	Moderate	Moderate	Non
>5.1	Moderate	Non	Non

DAS28 -disease activity score 28 joints. EULAR – European League Against Rheumatism. From Stevenson et al, 2016.

Traditional treatment for RA has been conventional disease-modifying anti-rheumatic drugs (cDMARDs) which include sulfasalazine and methotrexate, as well as corticosteroids, analgesics and non-steroidal anti-inflammatory drugs. From the late 1990s, a group of drugs have been developed consisting of monoclonal antibodies and soluble receptors that specifically modify the disease process by blocking key protein messenger molecules such as cytokines or cells such as B-lymphocytes. These drugs have been labelled biologic disease-modifying anti-rheumatic drugs (bDMARDS) (Stevenson et al, 2016). Tumour Necrosis Factor alpha inhibitors (TNFi) are a sub-set of bDMARDs. At the time of writing five drugs were licensed in the UK for the treatment of patients with Rheumatoid Arthritis (RA) who do not show or maintain adequate response to non-biological (or conventional) disease modifying anti-rheumatic drugs (cDMARDs) (Stevenson et al, 2016). Adalimumab (ADA) and Etanercept (ETN) are two of the five licensed TNFis. They are the drugs which were used in the clinical trial from which the data was taken to develop the response signature subject of this study. Rituximab (RTX) is an anti-CD20 monoclonal antibody that depletes various pathophysiological subsets within the B-cell population (Porter et al, 2016). Rituximab is approved for use in patients who have not responded to TNFi drugs. It was shown to be non-inferior to TNFi drugs (ADA and ETN) in the treatment of biologic-naïve, sero-positive patients (Porter et al, 2016).

Figure 8-1 illustrates the position of bDMARDs within NICE recommendations for sequence of treatments for patients with severe RA (DAS28>5.1) (adapted from Stevenson et al, 2016, Figure 1).

Figure 8-1: Treatment sequence for patients with RA in the UK (DAS28>5.1)



bDMARD – biologic disease modifying anti-rheumatic drugs; cDMARD – conventional disease modifying anti-rheumatic drugs; DAS28 – disease activity score 28 points, MTX – Methotrexate ; RTX – rituximab ; TCZ –tocilizumab (interleukin 6 inhibitor). Adapted from Stevenson et al, 2016.

For newly diagnosed patients a combination of cDMARDs (generally methotrexate (MTX) plus another cDMARD) is recommended (NICE, 2018c). If a patient has severe symptoms (DAS28>5.1) and has not responded to two cDMARDs then bDMARDs may be prescribed (usually TNF inhibitors). If TNFi fail then rituximab and methotrexate would typically be prescribed (Stevenson et al, 2016, p3). Currently, in the UK, an empiric strategy of drug selection is used as patients try a sequence of drugs, trying another if they fail to respond or stop responding.

There is potential for a test-based strategy to improve response rates if a test were able to identify which patients would respond to a specific bDMARD (Russell et al, 2018). There is currently no such test available. The technology in this case study (which is at an early prototype stage) is a series of three transcriptomic signatures to predict response to two types of biologic drugs in RA using a sample of the patient's blood. The signatures involve approximately 23 genetic markers combined with either age or gender of the patient depending on the signature. The signatures were developed using whole genome sequencing of samples collected during a clinical trial into the non-inferiority of the treatments. Now the genetic markers have been identified the next step is to develop a panel testing only the identified markers and to explore whether the signatures can be validated using this alternative platform.

The objective of this case study is to assess the potential clinical and economic value of the signatures at different time points following the approach set out in Vallejo-Torres et al (2011). A stage-gate is a term used in research and development to represent an important decision point when it is appropriate to assess the current position in the development and decide whether the project should proceed (Pietszch et al, 2009). This study examines two stage-gates in the development process of the testing technology. The early-stage analysis considers the evidence which was available prior to the whole genome sequencing (WGS) and analysis of the ORBIT trial data which identified the signatures. The aim of this aspect of the study is to determine whether there appeared to be prima facie value in the potential test which would justify the research expenditure. The mid-stage analysis includes the evidence from the WGS and subsequent analysis. The aim of this aspect is to begin to evaluate the prima facie case for value set out in the early-stage analysis. Output from this second stage can also begin to inform evidence generation strategy and technology design.

8.2 Early-stage analysis

8.2.1 Introduction

The basic value proposition of this technology is that a test and treat strategy may lead to an increase in response rates to the relevant biologic drugs. In order for the tests to add *clinical value* they have to improve response rates. In order to add *economic value* they either have to improve response rates or provide savings through prescription of a lower cost drug which are sufficient to justify the cost of testing all patients in the population. As the respective cost of drugs varies by jurisdiction and over time, this early-stage analysis focuses on the potential additional response rates generated by the tests.

8.2.2 Method

The early-stage assessment comprised a clinical value assessment. The clinical value assessment used a decision tree to show the health impact of the tests in terms of potential improved response rate.

Population, Intervention, Comparator and Outcome measures (PICO) criteria for the decision problem are set out in Table 8-2.

Table 8-2: PICO criteria for simple health impact assessment

Population	Adults with moderate to severe and severe RA who have not responded or stopped responding to cDMARDs so are eligible to commence biologic DMARDs.
Intervention	Test all for response to TNFi (treatment A) and treat according to result of the test. If not predicted to respond A treat with rituximab (treatment B).
Comparators	Treat all with TNFi or rituximab.
Outcome measures	Response rate at 6 months (good response and moderate/good response modelled separately)

cDMARDs – conventional disease modifying anti-rheumatic drugs, RA – rheumatoid arthritis, TNFi – tumour necrosis factor inhibitors

Decision trees with a time horizon of six months were developed to compare a strategy of testing all patients and prescribing treatment A (in this case TNFi) if the test predicts that a patient will respond. Otherwise a patient is prescribed treatment B (in this case rituximab). The second and third branches of the tree show strategies of treating all patients with either treatment A or treatment B (Figure 8-2 for good response and Figure 8-4 for moderate/good response). The decision tree generates an expected response rate from all three strategies. The additional response generated by a test and treat strategy over the best of the empiric treatment strategies can be compared to determine the level of additional response at a range of levels of sensitivity and specificity. The Test and Treat Superiority (TTS) Plot was produced from the odds ratio between the two treatments as explained in chapter 7. The threshold plotted on the TTS plot showed the combinations of sensitivity and specificity at which the testing strategy outperforms the best treat all strategy. Table 8-3 summarises the methods and data sources used in the early-stage assessment.

Table 8-3: Methods and data sources used in early-stage assessment

Method	Data	Source
Epidemiological analysis and Health impact assessment	Clinical pathway in UK Response rates to bDMARDs (see Table 8-4)	NICE CG100 (NICE, 2018c) SIGN guideline 123 (SIGN, 2011), Stevenson et al, 2016 ORBIT trial (Porter et al, 2016)

DMARDs – disease modifying anti-rheumatic drugs, HTA – health technology assessment, NICE – National Institute for Health and Care Excellence, SIGN – Scottish Intercollegiate Guidelines Network, TNFi – tumour necrosis factor inhibitors

Table 8-4 sets out selected response rates from a recent HTA report meta-analysis. It is evident that there are a proportion of patients who do not respond to the various drugs prescribed. Equivalent response rates for rituximab are given in the ORBIT trial as 0.29 for good response and 0.83 for moderate/good

response (Porter et al, 2016) although this is based on an intention to treat analysis so response rates will include patients who switch to the other treatment.

Table 8-4: Selected response rates taken from meta-analysis

Intervention	At least moderate response (95% CrI)	At least good response (95% CrI)
cDMARDs	0.410 (0.344 to 0.479)	0.077 (0.048 to 0.117)
ADA +MTX	0.664 (0.327 to 0.903)	0.220 (0.048 to 0.546)
ADA	0.704 (0.321 to 0.948)	0.254 (0.047 to 0.669)
ETN +MTX	0.871 (0.437 to 0.992)	0.473 (0.085 to 0.886)
ETN	0.670 (0.132 to 0.973)	0.224 (0.010 to 0.772)
Grouped biologics	0.711 (0.217 to 0.967)	0.260 (0.023 to 0.743)

ADA – Adalimumab, cDMARDs – conventional disease modifying anti-rheumatic drugs, CrI – credible interval - ETN – Etanercept, MTX – methotrexate. Taken from Stevenson et al, 2016.

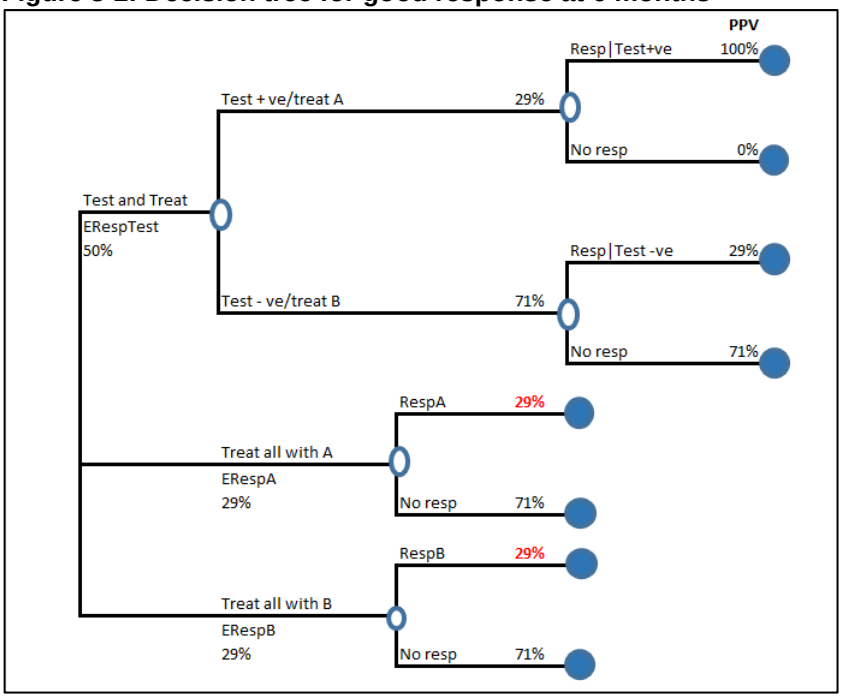
Pathways and data sources were based on the UK. This is because good quality evidence was readily available for the UK and the development was based here.

8.2.3 Results

8.2.3.1 Clinical value assessment

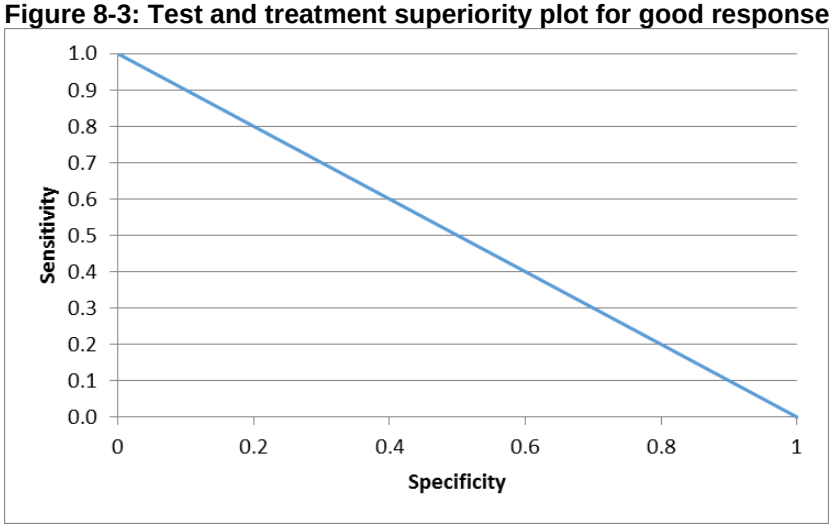
Set out below are the decision trees, additional response tables and TTS plots for good response and moderate/good response. Response rates are taken from Porter et al, 2016. Treatment A in this case is TNFi (so the test is for response to TNFi). The odds ratio of good response at six months, TNFi to rituximab was 1 as both treatments had response rates of 29%. For moderate and good response the odds ratio TNFi to rituximab was 0.65 - ((0.76/0.24)/(0.83/0.17)) reflecting the higher response rate achieved by rituximab of 83% compared to 76% for TNFi.

Figure 8-2: Decision tree for good response at 6 months



ErespA(B) – expected response from treatment A(B), ErespTest – expected response from test and treat strategy, resp – response, PPV – positive predictive value, Resp|Test+ve(-ve) – response rate conditional on positive (or negative) test result, Treat A – TNFi, Treat B, rituximab

Figure 8-3: Test and treatment superiority plot for good response at 6 months

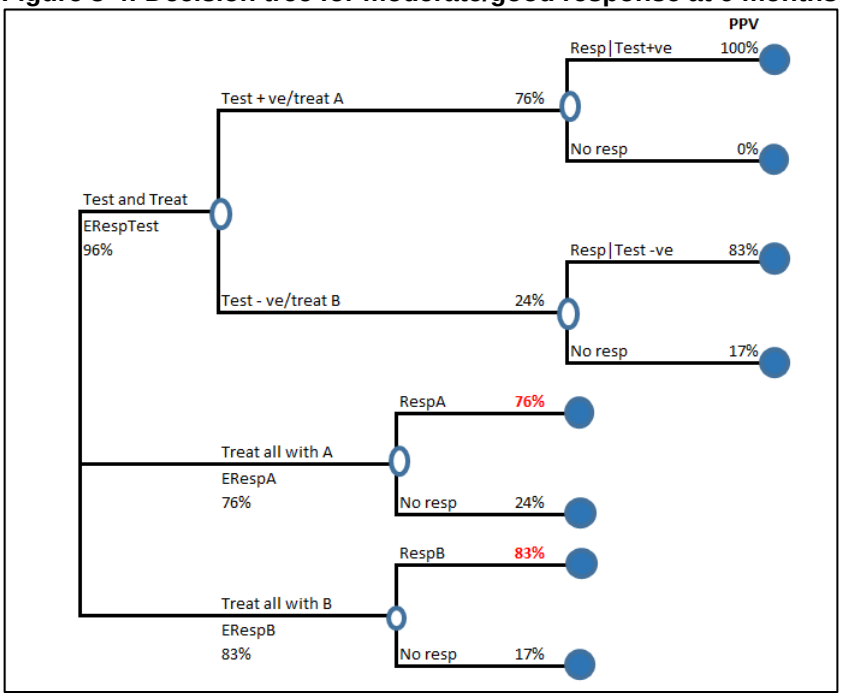


Above and to the right of the threshold line (Figure 8-3) indicates the area of test performance where the test and treat strategy outperforms the best of the treat all strategies. An odds ratio of 1 produces the threshold shown in Figure 8-3 where, because the treatments have an equal response rate as long as the test has a better than 50/50 performance it will add value.

Table 8-5: Response for test and treat strategy – good response

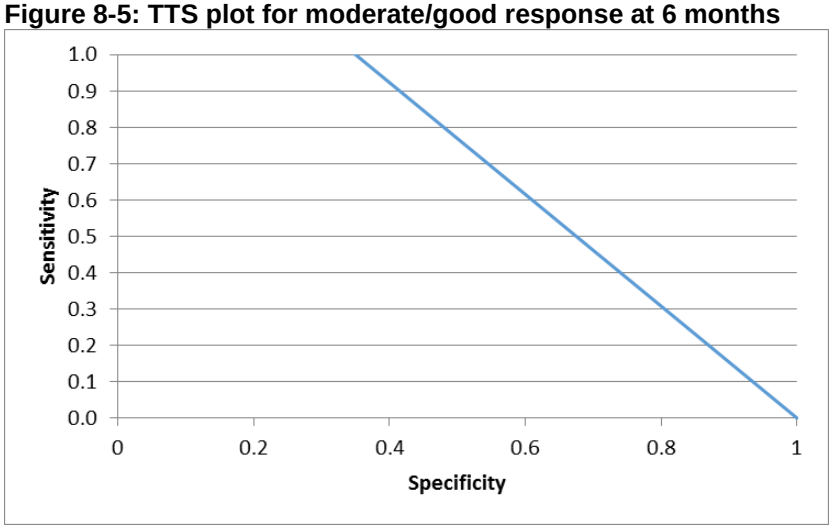
	Specificity	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Sensitivity	1	0%	2%	4%	6%	8%	10%	12%	14%	16%	19%	21%
	0.9	0%	0%	2%	4%	6%	8%	10%	12%	14%	16%	19%
	0.8	0%	0%	0%	2%	4%	6%	8%	10%	12%	14%	16%
	0.7	0%	0%	0%	0%	2%	4%	6%	8%	10%	12%	14%
	0.6	0%	0%	0%	0%	0%	2%	4%	6%	8%	10%	12%
	0.5	0%	0%	0%	0%	0%	0%	2%	4%	6%	8%	10%
	0.4	0%	0%	0%	0%	0%	0%	0%	2%	4%	6%	8%
	0.3	0%	0%	0%	0%	0%	0%	0%	0%	2%	4%	6%
	0.2	0%	0%	0%	0%	0%	0%	0%	0%	0%	2%	4%
	0.1	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	2%
	0	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%

Figure 8-4: Decision tree for moderate/good response at 6 months



EespA(B) – expected response from treatment A(B), EespTest – expected response from test and treat strategy, PPV – positive predictive value, resp – response, Resp|Test+ve(-ve) – response rate conditional on positive (or negative) test result, Treat A – TNFi, Treat B, rituximab

Figure 8-5: TTS plot for moderate/good response at 6 months



TTS – test and treat superiority

Table 8-6: Response for test and treat strategy – mod/good response

	Specificity										
	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Sensitivity 1	0%	0%	0%	0%	1%	3%	5%	7%	9%	11%	13%
0.9	0%	0%	0%	0%	0%	2%	4%	6%	8%	10%	12%
0.8	0%	0%	0%	0%	0%	0%	2%	4%	6%	8%	10%
0.7	0%	0%	0%	0%	0%	0%	1%	3%	5%	7%	9%
0.6	0%	0%	0%	0%	0%	0%	0%	2%	4%	6%	8%
0.5	0%	0%	0%	0%	0%	0%	0%	0%	2%	4%	6%
0.4	0%	0%	0%	0%	0%	0%	0%	0%	1%	3%	5%
0.3	0%	0%	0%	0%	0%	0%	0%	0%	0%	2%	4%
0.2	0%	0%	0%	0%	0%	0%	0%	0%	0%	1%	3%
0.1	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1%
0	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%

The above plots and tables demonstrate that a test and treat strategy could add clinical value provided that test performance exceeded the thresholds shown in Figure 8-3 for good response and

Figure 8-5 for moderate and good response. Additional good response of 21% (29% to 50%) and moderate to good response of 13% (83% to 96%) could be achieved. The TTS Plot shows equanimity between sensitivity and specificity for good response when the treatment response rates are equal (OR=1)(Figure 8-3) but specificity maximised when looking at moderate and good response as RTX (treatment B) has higher rates of response (

Figure 8-5). Three further points are worthy of note:

- In the UK, rituximab (RTX) is licensed for use after the failure of a TNFi. It is, therefore, clinically appropriate to test for response to TNFi rather than RTX. If sensitivity of the test is maximised this will optimise the identification of non-responders to TNFi. Clinicians may be more inclined to prescribe off-label if they had evidence that the patient would be unlikely to respond. Outside of the UK, the licensing position and clinician attitudes may differ.

- This simple modelling assumes that response rates are independent but this may not be the case. If there is overlap in the sense that it is the same patients who would respond to both treatments then this model will over-estimate the benefits of a test in increasing response rates. However, the test may still have economic value if it allows a cheaper treatment to be prescribed where a patient is likely to respond to both.

- The population in the ORBIT trial (Porter et al, 2016) were seropositive and their response rates may not be generalizable to the whole RA population.

8.2.4 Conclusion of early-stage analysis

The clinical value assessment demonstrates that a test of response to TNFi has the potential to increase response rates in the short term thus demonstrating potential clinical utility sufficient to justify initial investment in the development of the transcriptomic signatures from the ORBIT (Porter et al, 2016) data. Economic value of any test developed would derive partly from an ability to generate higher utilities in the period when response rates are higher than they would be under an empiric treatment selection strategy. The other aspect of economic value would be context specific and would depend on an ability to prescribe a cheaper treatment than TNFi if the test indicated that a patient was unlikely to respond. This aspect of value has not been calculated as the continued development of the test appears to be justified on the clinical value alone and this aspect is more generalizable.

A more detailed analysis follows which takes account of a greater variety of development scenarios and crucially incorporates the data generated from retrospective analysis of samples taken from responders and non-responders to both classes of drugs in the ORBIT trial (Porter et al, 2016). This provides an opportunity to include test accuracy in the models as well as different levels of overlap in response to treatments.

8.3 Mid-stage analysis

8.3.1 Introduction

The previous section set out how to articulate a basic value proposition for this technology in the absence of technology specific evidence. The second iteration will begin to evaluate the value proposition using the first technology specific evidence (from the retrospective analysis of the ORBIT trial data).

8.3.2 Method

The additional evidence incorporated in this stage of analysis comprised a retrospective analysis (unpublished) of samples collected during the ORBIT trial (Porter et al, 2016). The aim of the retrospective analysis was to identify transcriptomic signatures in the ORBIT cohort that can predict response/non-response to biologic therapy. The study comprised the Ribonucleic Acid (RNA) sequencing of 241 samples of which 70% (n=169) were used to develop models and 30% (n=72) used to validate the models. Clinical response was defined as a fall in DAS28-ESR of >1.2 units between baseline and 3 months. Multiple machine learning techniques were used to train models that predicted 1) general responsiveness to both TNFi and rituximab therapy and 2) differential response to TNFi or rituximab therapy. Three gene sets were identified that predicted general responsiveness to both TNFi and rituximab therapy (8 genes), response to TNFi (23 genes) and response to rituximab (23 genes). When tested on the validation set these models resulted in receiver operator curve (ROC) plots with an area under the curve of 91.6% for general responsiveness, 89.7% for TNFi and 85.7% for rituximab response. Patients who were predicted to respond at three months were also more likely to have a DAS28-good response (43% v 23%) or

DAS28 remission (23% v 10%) at 12 months (retrospective analysis of ORBIT data, unpublished).

A one-year decision-tree structure was developed in Microsoft Excel to explore the potential impact on response rates, cost and utilities of a test and treat strategy. The only costs included were assumed cost of test and the cost of drug treatment (including administration costs). The PICO criteria for the decision problem are set out in Table 8-7. Discounting was not applied as this is not appropriate for a 12-month time horizon.

Table 8-7: PICO criteria for cost-utility analysis

Population	Adults with moderate to severe and severe RA who have not responded or stopped responding to cDMARDs so are eligible to commence biologic DMARDs.
Intervention	Test all for response to TNFi and rituximab and treat according to result of the test. If predicted to respond to both or neither treat with TNFi.
Comparator	Treat all with TNFi.
Outcome measures	Response rates at 3 and 12 months. Response is defined as a reduction in more than 1.2 points on the Disease Activity Scale 28 (DAS28), 12-month costs and utilities and ICER.

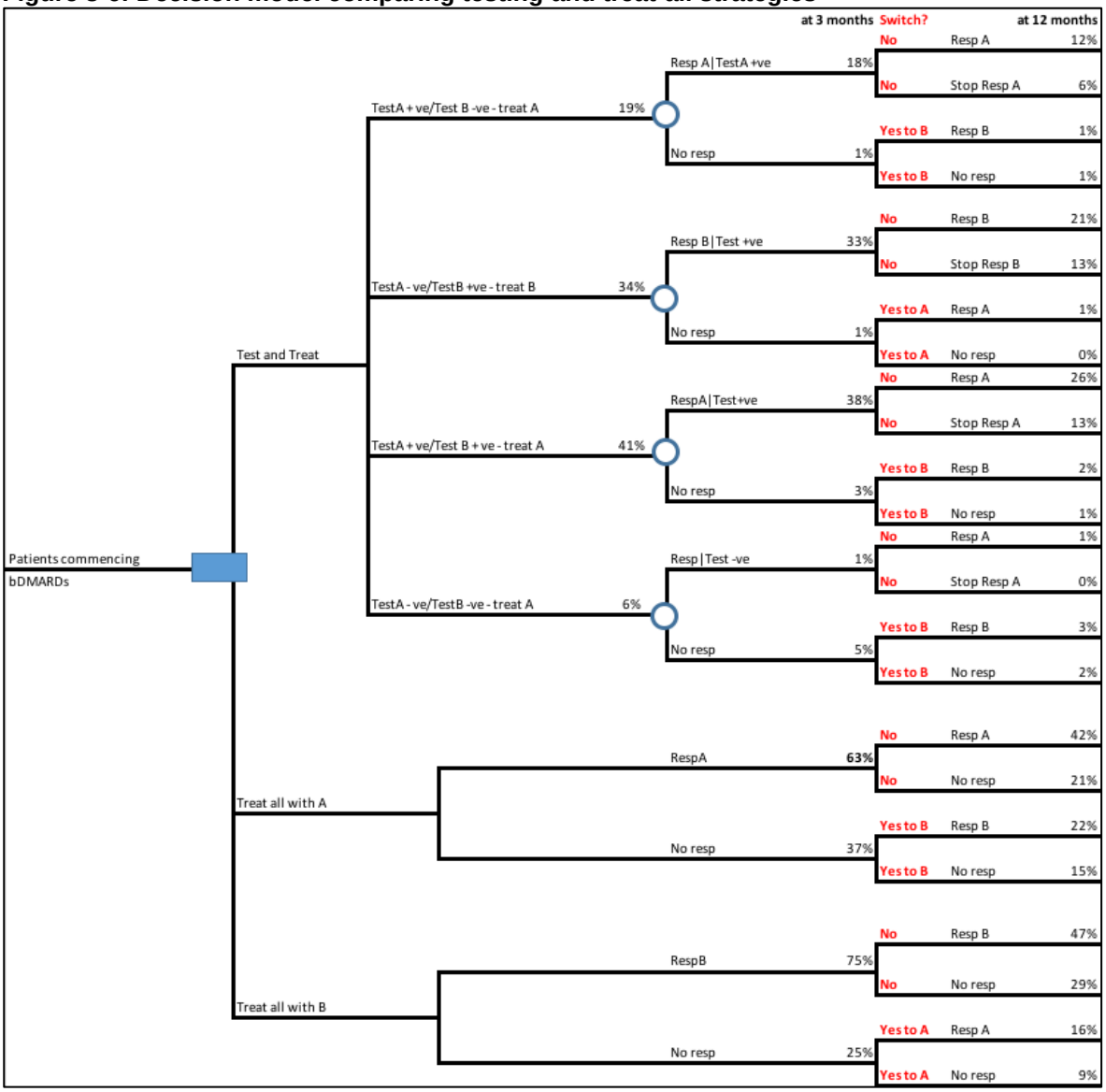
cDMARDs – conventional disease modifying anti-rheumatic drugs, ICER – incremental cost effectiveness ratio, RA – rheumatoid arthritis, TNFi – tumour necrosis factor inhibitors

The decision-tree presented in Figure 8-6 compares a test and treat strategy with strategies of treating all patients initially with TNFi (treatment A) and rituximab (treatment B). The test and treat strategy includes four options depending upon the result of the test. Where the test indicates that the patient is likely to respond to treatment A but not treatment B (Test A +ve/Test B -ve) the patient receives treatment A for the first three months. If the test indicates that the patient is likely to respond to treatment B but not treatment A (Test A --ve/Test B +ve) the patient receives treatment B for the first three months. If the test indicates that the patient is likely to respond to both treatments the base case assumes treatment A is given (as this corresponds to clinical guidelines and current practice in the UK). Where the test indicates that the patient is likely to respond to neither treatment the base case assumes that treatment A is still given. This reflects the fact that there are a number of alternative biologic modes of action available and no biologic treatment is unlikely to be acceptable to either clinician or patient. After the initial three months patients have responded or not. If they respond to the treatment received they remain on the treatment until 12 months but a proportion of patients will stop responding. If they do not respond they switch treatments at 3 months. They may or may not respond to the new treatment by 12 months. These periods were chosen as the

test is currently designed as a prediction of response overall (i.e. moderate/good response) at 3 months and 12 months is the latest data available in the dataset from the ORBIT trial. In practice, patients with severe disease are likely to be assessed every three months. The initial treatment may be retained longer than 3 months and treatment switching would occur between 3 and 12 months.

Inputs to the model and their sources are set out in Table 8-8. The key difference between this analysis and the simpler analysis set out in Section 8.2 is the availability of the data from the unpublished retrospective analysis of the ORBIT trial. In particular, this allowed for the estimation of test performance and the inclusion of estimated overlap of response to treatments. The column headed 'Range' gives the lower and upper limits used in the one-way sensitivity analysis. For response and test performance parameters these are the 95% confidence intervals. The upper limit for response at 9 months for switchers to rituximab was capped at 75% in line with overall response. This was necessary as confidence intervals were very wide due to small numbers and the calculated upper limit was over 90% response which was not consistent with the 3-month response rate. Observed persistence reflects the number of patients who responded to either treatment at 3 months and were still responding at 12 months. This was derived using survival analysis in Stata and confidence intervals are provided within the output from the analysis. The initial cost of testing was assumed to be £500. Costs of testing were varied between £250 and £1,000 (half and double the base case). Costs of treatment were increased/reduced by 35% as this is the estimated cost reduction associated with a biosimilar product (Manova et al, 2018). The utility values were based on those given in Stevenson et al, 2016. The HTA report gave values for responses for good, moderate and non-responders to TNFi. The value for non-responders is taken directly from Stevenson et al, 2016. The value for responders is a weighted average of good and moderate responders. Utility values are assumed to be the same whether the response is to TNFi or rituximab.

Figure 8-6: Decision model comparing testing and treat-all strategies



bDMARDs – biologic disease modifying anti-rheumatic drugs, No resp – no response to treatment, Resp A – respond to treatment A, RespB – respond to treatment B, TestA – test for response to TNFis, TestB – test for response to rituximab, +ve – positive result, patient likely to respond, -ve – negative result, patient unlikely to respond,

Table 8-8: Inputs to cost-utility model

Input parameter		Base case	Range	Source
Patients initiating biologic DMARDs annual - England and V	cases	6,000		Stevenson et al, 2016
Patients predicted to respond to TNFi and Rituximab	predR_overlap	41%	34% -47%	Reanalysis of ORBIT trial data (unpublished)
Sensitivity of test for response to TNFi	sensA	89%	81%-94%	Reanalysis of ORBIT trial data (unpublished)
Specificity of test for response to TNFi	specA	89%	81%-94%	Reanalysis of ORBIT trial data (unpublished)
Sensitivity of test for response to Rituximab	sensB	98%	93%-100%	Reanalysis of ORBIT trial data (unpublished)
Specificity of test for response to Rituximab	specB	93%	86%-97%	Reanalysis of ORBIT trial data (unpublished)
Response to TNFi at 3 months	pA	63%	53%-72%	Reanalysis of ORBIT trial data (unpublished)
Response to Rituximab at 3 months	pB	75%	65%-83%	Reanalysis of ORBIT trial data (unpublished)
Response for switchers to TNFi at 9 months	pA_sw	64%	35%-87%	Reanalysis of ORBIT trial data (unpublished)
Response for switchers to Rituximab at 9 months	pB_sw	60%	15%-75%	Reanalysis of ORBIT trial data (unpublished)
Observed persistence in TNFi response rate by 12 months	OP_A	67%	56%-76%	Reanalysis of ORBIT trial data (unpublished)
Observed persistence in Rituximab response rate by 12 m	OP_B	62%	51%-71%	Reanalysis of ORBIT trial data (unpublished)
Cost of test (£)	cost_test	500	250-1,000	Assumed
Cost of treatment with TNFi (etanercept)- 6 months (£)	cost_A	4,648	3,021-6,275	NICE Technology Assessment 195
Cost of treatment with Rituximab - 6 months (£)	cost_B	3,492	2,270-4,714	NICE Technology Assessment 195
Annual utility for responder	u_resp	0.52	0.42-0.61	Calculated based on Stevenson et al, 2016
Annual utility for non-responder	u_nonresp	0.31	0.19-0.42	Calculated based on Stevenson et al, 2016

8.3.3 Results

Table 8-9: Results of cost-utility analysis

Strategy	3-month response	12-month response	Utility	Costs (£)	ICER (£)
Test and treat	90%	65%	0.4581	8,867.01	n/a
Treat all TNFi	63%	64%	0.4445	8,654.42	n/a
Incremental test and treat vs treat all TNFi – base case analysis	27%	0%	0.0136	212.59	15,613
Treat all rituximab	75%	63%	0.4478	7,417.50	n/a
Incremental test and treat vs treat all rituximab	15%	2%	0.0103	1,449.51	140,393

The results of the base-case analysis show an improvement in response rate at 3 months of 27% from 63% response under a treat all with TNFi. By 12 months no additional response is generated by the testing strategy as over this time-frame all patients will have had the opportunity to try both modes of action. The difference in response in the first three months leads to an incremental utility of 0.0136 at an incremental cost of £212.59. The ICER at £15,613 is below the lower threshold set in the UK for reimbursement of health technologies of £20,000. This demonstrates that, if the assumptions in the base case held (including, in particular the assumption that the developers were able to offer the test for £500) then the technology may be eligible for reimbursement in the UK.

The ICER for the test and treat strategy compared to treat all with rituximab is considerably higher than when the test and treat strategy is compared with treat all with TNFi. This is because the response rate to rituximab is higher in the first three months than the response to TNFi so the incremental gain of moving some patients to TNFi is not as high as moving patients in the other direction. Additionally, rituximab is cheaper than TNFi so any patients moved to TNFi will incur more cost. The comparison against treat all with TNFi is given as the base case analysis as this more closely reflects clinical practice in the UK.

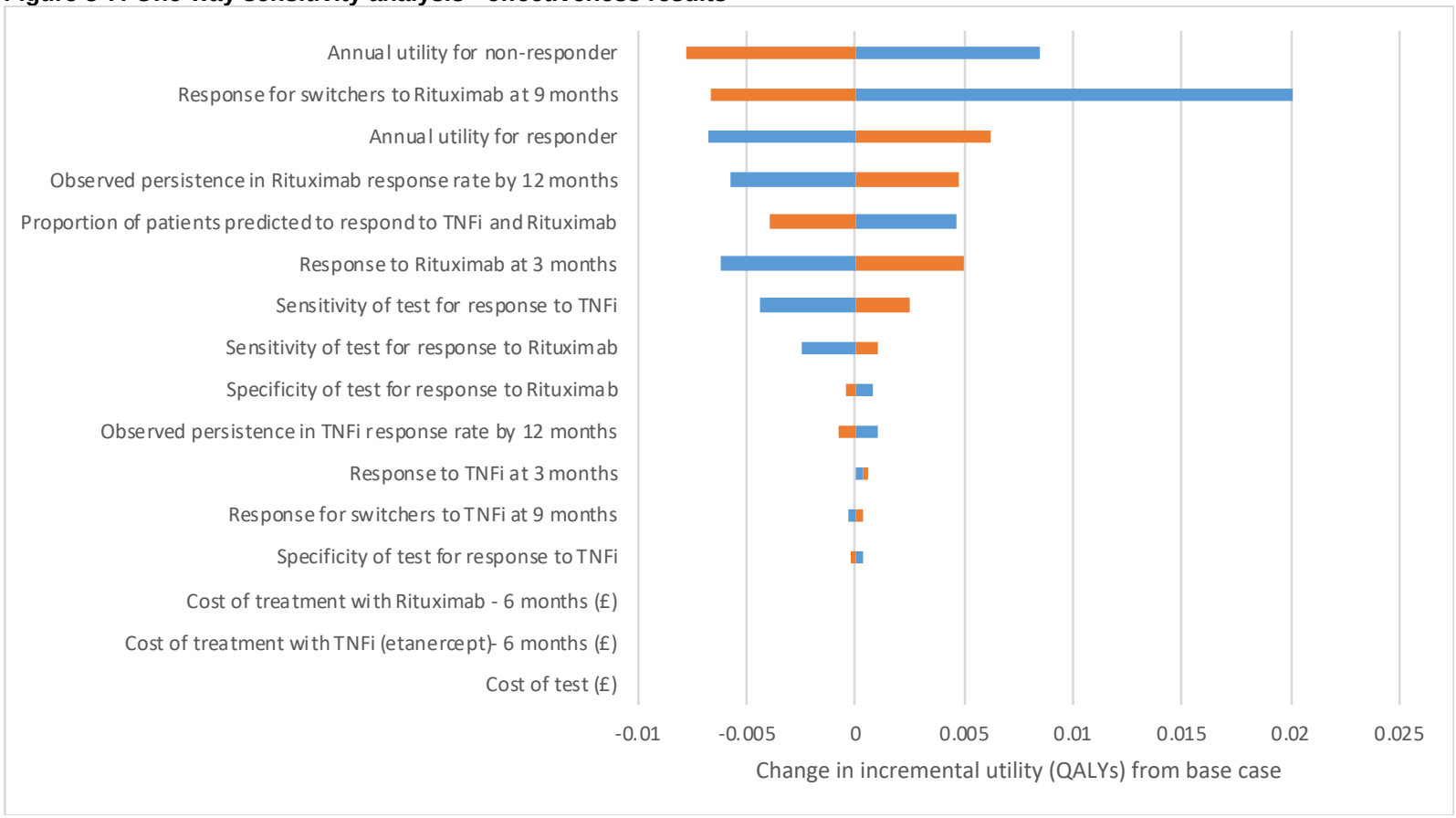
One-way sensitivity analysis was undertaken to assess the impact of each input parameter on incremental clinical effectiveness, incremental costs and the ICER. The analysis shows that the effectiveness result (Figure 8-7) is most sensitive to utilities for response and non-response, rituximab response rates at 3 months

and at 9 months for switchers, predicted overlap in response between the treatments and observed persistence in response to rituximab at 12 months. The effectiveness results are less sensitive to test performance parameters although they are more sensitive to sensitivity than to specificity of tests for both treatments.

The incremental cost results are most sensitive to the cost of the treatments and the cost of the test. As the test cost is reduced to the lower bound (£250) the ICER is negative as the testing strategy reduces costs and improves clinical effectiveness. These results reflect the cost of each treatment moving in isolation (i.e. the price of TNFi reducing but rituximab staying the same). Where TNFi prices reduce this results in a lower saving so increases the ICER. This is because part of the benefit of the testing strategy comes from moving patients onto a cheaper treatment. The opposite is true for rituximab as the price decreases this improves the saving and results in a negative ICER due to an overall cost saving of the testing strategy. Relative and absolute costs of the treatments vary depending upon the setting of use of the test and over time and will be a critical factor for the developers to consider when selecting markets to prioritise. The introduction of biosimilars will also impact upon pricing - an issue that is addressed in Scenario 1. The cost of the test is, as yet, unknown and costs are changing rapidly as platforms and processes develop. Again, this will be a critical factor for the development team to monitor.

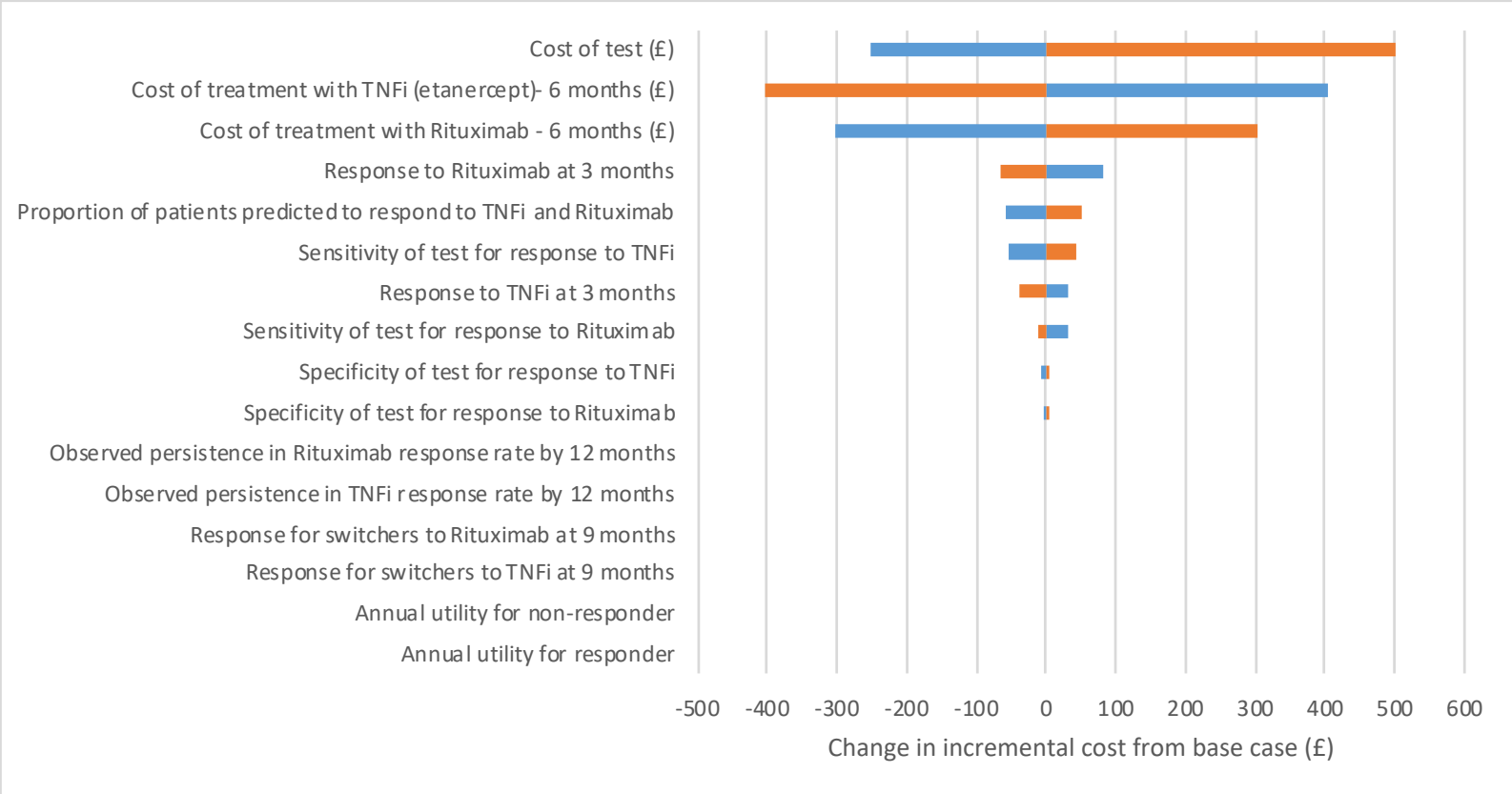
Figure 8-9 shows that the ICER is most sensitive to the cost parameters, cost of the test and the costs of both treatments. The ICER is moderately sensitive to rituximab response parameters, the overlap of patients predicted to respond to both treatments and longer term (9 month) and utility parameters. The ICER is not sensitive to any of the test performance parameters within the ranges tested.

Figure 8-7: One-way sensitivity analysis - effectiveness results



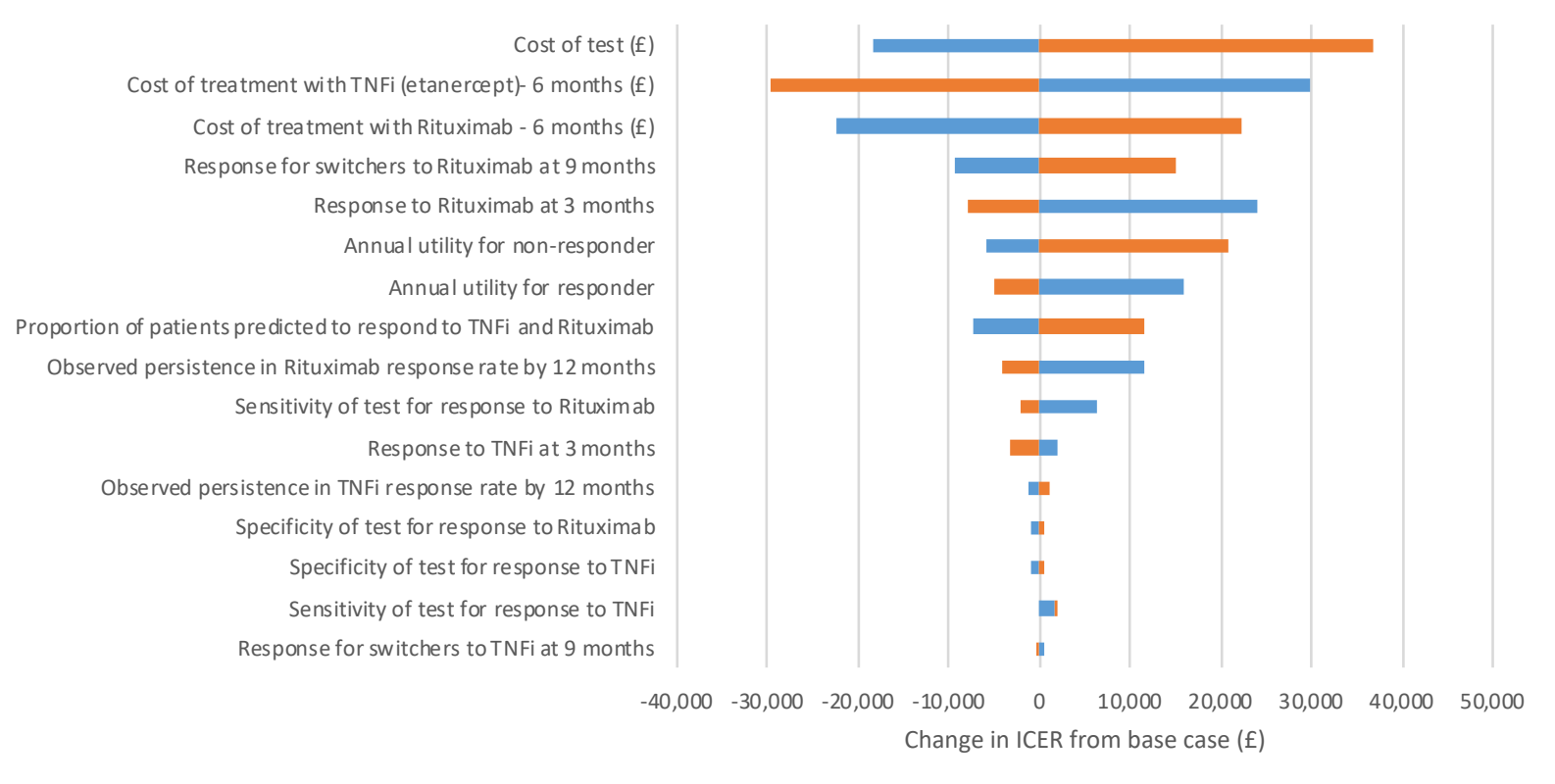
Red effect of lower limit, blue effect of upper limit

Figure 8-8: One-way sensitivity analysis - costs results



Red effect of lower limit, blue effect of upper limit

Figure 8-9: One-way sensitivity analysis – ICER



Red effect of lower limit, blue effect of upper limit

Scenario analysis involves exploring alternative plausible scenarios which could impact on the value of the technology in development. One scenario explores the impact of biosimilars, three concern response rates and the extent of the overlap between response to TNFi and rituximab and a final scenario addresses behavioural aspects of use of test results.

The introduction of biosimilars will change the potential economic impact of a test and treat strategy. Assuming that the tests in development were equally effective in identifying responders to biosimilars as in the drugs they are designed to copy, pricing is very different. It is, therefore, appropriate to model scenarios where a biosimilar displaces either or both TNFi and rituximab. The impact modelled will be confined to modelling price change (i.e. assuming that market penetration of the test would be unaffected by the market presence of biosimilars). One-way sensitivity analysis in the cost-utility section models the impact of a change in the price of either drug – scenario 1 below models the impact of a change in the price of both drugs simultaneously.

Another aspect of importance is the extent of the overlap between the responders to TNFi and rituximab. This overlap can always only be estimated as by definition patients can only be prescribed one treatment as their initial biologic treatment. The base case analysis takes the extent of the overlap in response rates from the predictions made using the transcriptomic signature in development. Scenario 2 tests the impact of using equal response rates (the base case uses the response rates from the ORBIT trial (Porter et al, 2016) and takes the overlap from the retrospective analysis of the ORBIT trial data (as in the base case). Scenarios 3 and 4 retain equal response rates but assume a complete overlap in responses and minimum overlap in responses, respectively (see Figure 8-10).

A further risk is that clinicians and/or patients may be reluctant to apply the results of the diagnostic test (Thompson et al, 2014). Scenario 5 models the impact of 25% of decisions not following the recommendation of the test. In effect this means that 25% of patients are treated empirically under the test and treat strategy.

Table 8-10: Results of scenario analysis

Strategy	Change in assumptions	Incremental response rates at 3 months, 12 months, incremental utility, incremental costs and ICERs				
		3-month response	12-month response	Utility	Costs (£)	ICER (£)
Base case	See Table 8-8	27%	0%	0.0136	212.59	15,613
Scenario 1	Biosimilars replace TNFi and rituximab with 35% price reduction across the board	27%	0%	0.0136	313.28	23,008
Scenario 2	Equal response parameters to TNFi and rituximab and test performance – overlap in responders to TNFi and rituximab as in base case	16%	-1%	0.0035	336.89	97,395
Scenario 3	As Scenario 2 but minimum overlap	25%	-1%	0.0097	251.45	25,931
Scenario 4	As Scenario 2 but full overlap	-2%	0%	-0.0102	523.30	-51,550
Scenario 5	25% of patients continue to be treated empirically under test and treat strategy	17%	0%	0.0092	266.02	28,823

Scenario 1 – price reduction for treatments

As shown in Figure 8-10, one-way sensitivity analysis shows that If the cost of rituximab reduces by 35% with all other parameters held equal the overall test and treat strategy dominates the existing strategy as it is cheaper and more effective. This is because a greater proportion of patients are treated with rituximab for longer using the test and treat strategy. If the cost of rituximab goes up by 35% the test and treat strategy is no longer cost effective at UK threshold levels (ICER £37,925). This is because, under base case assumptions, the cost of testing is partially offset by a greater proportion of patients being treated with the cheaper treatment. With a 35% increase, rituximab is more expensive than TNFi and the improvement in utilities achieved through the improved response rates is not sufficient to justify the increased cost of universal testing. If the price of TNFi reduces by 35%, while all other parameters remain fixed, including the price of rituximab, the test and treat strategy is no longer cost effective under UK thresholds as there are no treatment cost savings to offset the cost of testing. If the cost of TNFi increases by 35% then the test and treat strategy is cost saving and dominates the treat all strategy (as when the cost of rituximab reduced). It is clear that the relative pricing of the two drugs is key to the cost-effectiveness of the test in the United

Kingdom and other jurisdictions that use similar thresholds to determine whether the technology will be made available.

Scenario 1 extends the one-way sensitivity analysis to show what the impact is of both treatments reducing by 35%. This scenario is possible as biosimilars replace both TNFis and rituximab (Manova et al, 2018). From, it can be seen that the ICER increases in the event that the prices of both treatments drop. This somewhat counter-intuitive result is because the cost savings from treating a greater proportion of patients with the cheaper treatment are not so great in absolute terms. As the cost of testing remains unchanged the test and treat strategy becomes less cost-effective.

Scenario 2 – Equal response rates to TNFi and rituximab holding overlap constant at 41%

The purpose of this scenario is to investigate the extent to which the base-case result depends on the higher response rate for rituximab suggested by the ORBIT trial reanalysis. In this scenario, the proportion predicted to respond to both treatments is held constant at 41% (as in the base case) but the overall response rates are equalised at 65% (24% responding only to TNFi or rituximab and 41% to both). This compares with overall response rates of 75% for rituximab and 63% for TNFi. Similarly, test performance, observed persistence rates and response rates for switchers are also equalised at 90%, 60% and 60% respectively. The results of this analysis show that the test and treat strategy still increases the overall response rate in the first 3 months by 16% (compared to 27% under the base case). However, the impact of this reduced incremental response rate on the utility gain and higher costs due to a higher proportion of patients on TNFi results in an ICER of £83,191.

Scenario 3 – Equal response rates – minimum overlap

Retaining response rates for TNFi and rituximab equal at 65%, this scenario models the impact of the testing strategy if there is the least possible overlap in responders. For response rates of 65% this minimum overlap is 30% (compared to the base case of 41%). Scenario 3 results in the testing strategy increasing response rates at 3 months by 25% over the treat all strategy. The gain in utility

is 0.0116 at an incremental cost of £338.56 resulting in an ICER of £29,074 compared to the treat all strategy. This improvement over the treat all strategy is possible because the testing strategy allows a greater proportion of patients to be prescribed their optimum treatment in the first three months. It is not as favourable as the base case due to the equalisation of response rates closer to TNFi response rates.

Scenario 4 -Equal response rates – full overlap

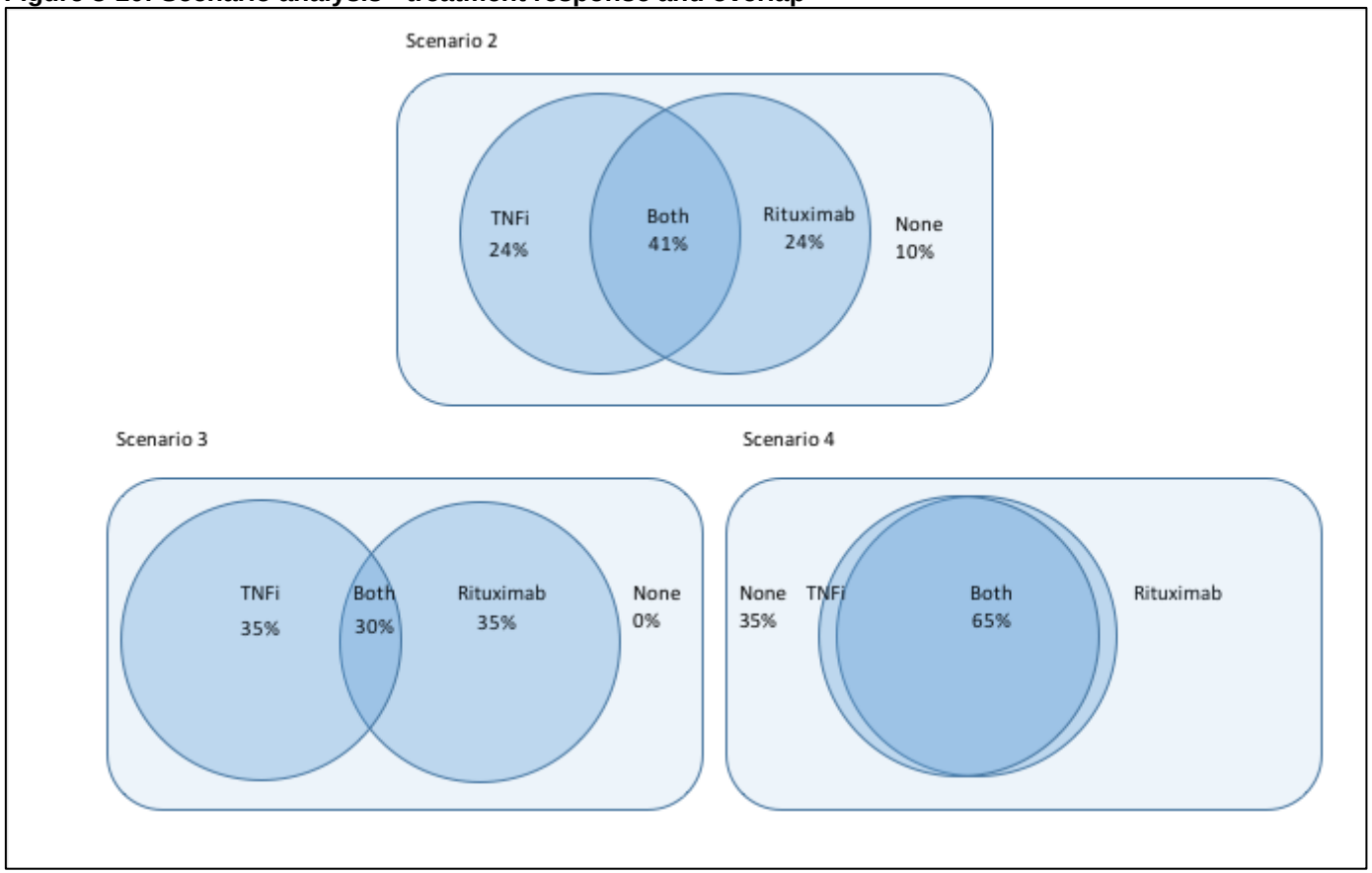
Scenario 4 results in the testing strategy reducing response rates at 3 months by 2% compared to a treat all strategy. This is because the tests are not perfect and response rates to the two treatments are equal. Thus, there is no opportunity to improve response rates by treating with one treatment compared to another. This reduction in response rates results in reduced utility and testing increases costs over a treat all strategy so that the testing strategy is dominated by the strategy of treating all with TNFi.

The model makes the assumption that patients who will respond to both treatments are prescribed TNFi in the first instance, as this follows licensing and clinical guidance in the UK. However, where it is acceptable to prescribe either treatment, there is an opportunity for the cheaper treatment to be prescribed when responses overlap. The bigger the overlap, the larger the likely cost savings to be made.

Scenario 5 – 25% of patients continue to be treated empirically under test and treat strategy

The purpose of this scenario is to evaluate the impact of behavioural aspects known to affect diagnostic technologies. The scenario assumes that for 25% of patients who are tested the clinician and/or patient do not follow the result of the test but prescribe as if under the treat all strategy. At 25% test and ignore, the response rate improvement at 3 months is reduced to 17% which results in an ICER just below the £30,000 upper threshold in the UK. It will be key for the developers to investigate which factors may lead clinicians and/or patients to test but then disregard the results and prescribe empirically.

Figure 8-10: Scenario analysis - treatment response and overlap



Scenario 2 – response rates equal, non-response rate and overlap equivalent to base case; Scenario 3 – response rates held at 65% but overlap reduced to minimum without overall response rate exceeding 100%; Scenario 4 – response rates held at 65% but complete overlap of responders assumed so that non-response rate increases to 35%

8.3.4 Conclusion of mid-stage analysis

The updated clinical value assessment and the economic value assessment demonstrate that a test of response to TNFi has the potential to be cost-effective in a UK setting but that the cost-effectiveness is particularly sensitive to the relative cost of the treatments under consideration. The analysis supports the continuation of the development at this stage-gate subject to a number of considerations discussed in the general conclusions for the case study in section 8.4.2

8.4 Discussion

8.4.1 Strengths and limitations

This case study demonstrates several strengths of the methods of development-focused HTA. Firstly, the simple modelling provides a means of communication between the development team and other stakeholders and facilitates discussion around the technology under development. Rheumatoid arthritis is a complex disease to model and full models are both resource-intensive and difficult to interpret. This model allows the development team and other stakeholders to see the impact on short term response rates of the different parameters – and the cost associated with achieving those increases. The complex aspects are not ignored but are dealt with outside of the modelling. As well as showing the impact of the different parameters on the response rates the model allows different scenarios to be assessed. The case study also demonstrates the value of the wealth of information available from a rapid literature review of key sources such as HTA reports and NICE and SIGN guidelines for the mapping of the clinical pathways and potential health impact of the technology. The iterative approach showed that analysis can be undertaken before any technology specific evidence is available thus providing a prima facie case of clinical value before any research cost has been incurred.

Development-focused HTA incorporates methods with a wider remit than reimbursement-focused HTA and the inclusion of scenario analysis taking into account the dynamic therapeutic environment in RA ensures that issues such as the potential impact of biosimilars and JAK inhibitors are explicitly considered.

Furthermore, the modelling identified areas where clinician and patient output were required (such as preferred outcome measure and likelihood of adhering to result) suggesting that stakeholder consultation would be appropriate.

The approach adopted in this case study also had a number of limitations. The analysis assumed a class action for TNFi. This may not be the case and Stevenson et al (2016) did not make a class action assumption in their multiple technology assessment of bDMARDs. However, Stevenson et al (2016) did comment that there was little difference between most of their modelled strategies. Whilst it may be an acceptable approach at this stage, the identification of a prediction signature for a particular drug is likely to be more acceptable to clinicians and patients as well as more helpful in deciding between the increasing number of drugs offered in this area. Scenarios modelled to date all assume that TNFi is prescribed both for non-responders and for those patients predicted to respond to both TNFi and rituximab. Although this may under-estimate the potential benefit of the test and treat strategy it is likely to be representative of clinical practice as rituximab is not licensed for use before a TNFi in the United Kingdom and other Western jurisdictions. In markets where clinicians were prepared to prescribe rituximab before a TNFi and rituximab was cheaper, there may be potential to realise a cost-saving by prescribing rituximab to those who were predicted to respond to both although this is highly dependent upon the proliferation and pricing of the many biosimilars currently being introduced. The population in the ORBIT trial were all sero-positive. There is a possibility that estimates of response improvements are increased as response to rituximab is thought to be higher in sero-positive patients (Porter et al, 2016). The model assumed that utilities were equivalent across responders to different treatments and for all levels of response. An improvement to the model would be to consider other levels of response such as good response or remission over other time periods. A final limitation is that the fact that the analysis uses a UK-centric approach. Other jurisdictions have very different ways of assessing whether a technology is deemed cost-effective and there is potential to market directly to patients. It is key that the development team take a broad view about which jurisdiction in which the technology may be most viable and valuable.

8.4.2 Conclusions, implications and next steps for the development team

Conclusions from the development-focused HTA exercise in this case study can be linked back to the decisions that the exercise was seeking to inform which were technology design, evidence generation strategy and go/no go decision. Aspects of technology design that could be informed in this case study included required test performance, choice of outcome measure and placement of the tests in the diagnostic/treatment pathway. The HTA has shown that test performance is not critical to the value proposition of the tests but that treatment response rates and the overlap between the response to the different treatments are important. As more data are generated, some further investigation into alternative thresholds may be warranted to assess the impact of maximising sensitivity or specificity to either of the treatments. This may be particularly important if different measures of outcome are being considered.

Choice of outcome measure may be of key importance as the initial value proposition assessment demonstrated that if the odds ratio for the treatments is different for the quality of response, then this may result in a different value-based price. The data available meant that the prediction of response could only be modelled for response overall at three months, however, in later stages of development it would be informative to model different levels of response at different time points. Patients and clinicians may also have preferences for particular outcome measures (both in terms of quality of response and time of response) so this aspect may impact upon uptake of the test. Some initial qualitative work has been undertaken by the project team to begin to assess this but is not reported here. In other jurisdictions, utility may not be the method of assessing health outcome and it may be worth investing in either some direct elicitation of patients' willingness to pay for the technology in key jurisdictions or collecting data on willingness to pay for comparable technologies.

In terms of placement in the pathway, the current exercise modelled a plausible clinical scenario for the use of the tests in the United Kingdom. This involved testing for both TNFi and rituximab response and assuming that everyone is treated with TNFi apart from those patients predicted to respond to rituximab but not to TNFi. It is plausible that rituximab could be prescribed to the group

predicted to respond to both which may result in cost savings (depending on relevant price differentials). It may also be plausible to look at testing for response to both agents earlier in the treatment pathway in some jurisdictions if clinicians are prepared to prescribe bDMARDs to patients either before or with a shorter exposure to cDMARDs. The tests may have greater potential value used earlier and in conjunction with a test for non-response to cDMARDS. If such a test could identify non-responders to cDMARDs who would respond to bDMARDs then this could considerably accelerate treatment decisions and prevent structural damage and disutility (and other consequences such as loss of employment). This more aggressive treatment approach may be suitable in a younger age group where employers' insurance schemes may take productivity losses into account and be prepared to fund more expensive treatment regimens.

Evidence generation strategies was the second area that the development-focused HTA exercise aimed to inform. The one-way sensitivity analysis identified treatment response and costs of both the test and treatments as key parameters. Neither costs nor effectiveness were highly sensitive to test accuracy parameters. The next stage in development involves the validation of the transcriptomic signatures using a panel. If the signatures continue to show the same level of accuracy, then a further evidence generation stage will be required. The developers need to consider how to design this evidence gathering in the most efficient way. Prospective trials are not always required by regulators and early contact with relevant regulatory/scientific bodies to ensure that evidence generated would meet all regulatory requirements is recommended. Levin (2015) explains how this single, harmonised approach has been developed in the MaRS Excite programme in Ontario, Canada (MARS, 2020). NICE also offer a scientific advice service (NICE, 2018e) to advise developers about the kind of evidence that would be required to demonstrate the value of the technology. It may be possible to use other repositories of samples and observational data to extend the retrospective validation of the signatures to a wider range of bDMARDs and to include both sero-positive and sero-negative patients. It may also be valuable to consider likely price differentials and clinical guidelines in alternative jurisdictions in order to prioritise the test into those jurisdictions where the technology could be most valuable. A partnership

with one of the pharmaceutical companies active in the area may provide access to extensive global knowledge of these aspects.

The final decision which the exercise aimed to inform was a decision whether to continue to invest or a so-called go/no-go decision. At both the initial and the second stage the tests generated additional response to bDMARDs in the short term. This indication of clinical utility means that the technology is likely to have an economic value to some patients/clinicians in some jurisdictions. The cost-utility analysis suggested that the technology would meet the criteria to be made available in the UK NHS provided it was assessed as a companion diagnostic. However, the utility gain generated was quite small and the ICER was very sensitive to movements in the relative costs of the treatments and to the cost of the test itself. The cost of the test is currently unknown and difficult to estimate due to the speed of technological advancement. The relative cost of treatments is also uncertain and varies across jurisdictions due to patent expiry, licensing status of biosimilars and the emergence of new modes of action in this therapeutic area. Given that this technology is believed to be a 'first in class' i.e. no comparable technology exists it would seem reasonable to take an optimistic view of its potential and to advise that the development team continue to invest.

In summary, the development team may wish to consider:

- a wider (treatments and populations) retrospective analysis in large observational datasets if blood samples are available

- a partnership approach with one of the pharmaceutical companies with knowledge of global treatment pathways and costing structures

- contacting licensing/regulatory authorities to ensure next stage of evidence generation is acceptable for all purposes

- engaging with stakeholders to ensure that the test is acceptable to patients and clinicians and that design is optimised to maximise impact on response rates, use and adherence

- including behavioural aspects in any future evidence generation exercise

- to investigate opportunities to use the test earlier in the treatment pathway potentially in conjunction with a test of non-response to conventional DMARDs.

9 Case study 3 - HTA of extension to molecular pathology testing in colorectal cancer

9.1 Introduction

This chapter is based upon a manuscript accepted for publication in the International Journal of Technology Assessment in Health Care in May 2019. The co-authors are Yun Yi Tan and Janet Graham (clinicians), Gillian McGaffin, Graham Smith, Paul Westwood and Nicola Williams (pathologists), David Creed (employee of Merck Serono Ltd), Neil Hawkins (health economist) and Ruth McLaughlin (previously project manager of Glasgow Molecular Pathology Node). Yun Yi Tan undertook the original clinical audit edited the draft for clinical content. Neil Hawkins validated the model and read and approved the manuscript. Ruth McLaughlin was involved in the original design of the project. The remainder of the co-authors were members of the expert group and read, edited and approved the manuscript.

The developers for this case study are a team of clinicians and pathologists working in the NHS in Scotland. The study arose following a clinical audit of the pathway for molecular pathology results in metastatic colorectal cancer in the West of Scotland. Clinicians were concerned that due to organisational factors results of key molecular pathology tests were not known at the time of the patient's initial visit in a high proportion of cases. The clinicians felt that this was resulting in sub-optimal consultations with patients and leading to some treatment choices being driven by the need to keep certain treatment options open.

The **stance** of the HTA is normative in that there is a clear, proposed service change and are seeking to evaluate the change against a known set of criteria in operation within NHS Health Scotland. MPEP consider evidence about costs of tests individually and for the national population as well as any savings or investment in the diagnostic or treatment pathways resulting from the introduction of the new test (NHS National Services Division, 2018). The evidence is provided by the NHS staff applying for the extension to molecular pathology testing. There is no requirement for modelling and no resource provided to undertake this. The specific decision to inform is a one-off binary

decision, whether or not to extend the testing to a larger patient population. Although the logical **core decision rule** should be to allow the extension if it provides improved health outcomes for a given budget, it is unclear the extent to which the evaluation process provided for by MPEP facilitates a full evaluation. In particular, it is unclear whether the system enables proposers of extensions to testing to demonstrate the full effects of a change throughout the diagnostic and treatment pathways and on both health outcomes and costs. The evaluation process is not developed as it is for pharmaceuticals and companion diagnostics which are assessed by the Scottish Medicines Consortium (SMC) against published criteria (Scottish Medicine Consortium, 2018a). Medicines (and their companion diagnostics) are mandated if they are recommended by the SMC and they are likely to be recommended provided they have an incremental cost-effectiveness ratio below a known threshold. This means that an increased cost of testing can be mandated as long as there is a corresponding increase in health outcomes. For stand-alone diagnostics there is no parallel procedure. Formal health economic evaluation is not undertaken for non-companion diagnostics although cost implications are clearly taken into account in some manner. It is, therefore, possible that the core decision rule is more concerned with the impact of accepting the test on the budget for molecular pathology testing rather than on the cost-effectiveness of the change for the diagnostic and treatment pathway as a whole.

The absence of a health economic evaluation framework for stand-alone molecular testing means that there is no established burden of proof for this work and that there is little resource available for the analysis or for evidence gathering/generating activities. Thus, the similarity to development-focused work arises from the absence of a clear structure of evaluation for stand-alone molecular tests and the committed resource that accompanies such a structure.

Colorectal cancer (CRC) is second only to lung cancer in incidence in Scotland (ISD Scotland, 2018a). 3,700 cases were diagnosed in 2016 and it was the cause of 1,687 deaths in 2017 (ISD Scotland, 2018a). Cancer services in the National Health Service (NHS) in Scotland are organised on a regional basis with the West of Scotland region accounting for 1,682 (45%) of the incident cases in 2016 (ISD Scotland, 2018a). The West of Scotland Cancer Network (WoSCAN) serves a

population of 2.5 million people (approximately 46.5% of the Scottish population) (WOSCAN, 2018). Cancer care in the West of Scotland region is delivered at 15 hospitals funded by 4 regional health boards (WOSCAN, 2018) supported by a centralised molecular pathology laboratory funded on a national basis (NHS National Services Division, 2018). Molecular biomarker testing in CRC was standardised across Scotland in 2015 when the Molecular Pathology Steering Committee approved a national patient testing pathway whereby all patients with metastatic disease were offered testing for mutations in KRAS, NRAS (collectively referred to as RAS) genes, both of which encode proteins involved in the epidermal growth factor receptor (EGFR) pathway, and in the BRAF gene, which is a downstream effector of the RAS genes. The aim of this national pathway was to ensure both equity of service across Scotland and also effective patient stratification for prognostic or therapeutic purposes.

RAS and BRAF testing in the West of Scotland are currently carried out on request after the confirmation of metastatic disease and through the multi-disciplinary meeting (MDM) where clinical colleagues discuss a patient's results and treatment. This is believed to be the current practice in most NHS settings (NICE, 2018b). RAS mutation status impacts upon treatment options as only patients with RAS wild-type disease (i.e. no NRAS or KRAS mutations) would be offered epidermal growth factor receptor inhibitors (EGFRi) - such as panitumumab and cetuximab - and these drugs must be given in conjunction with an infusional intravenous (IVI) chemotherapy regimen including a 48-hour 5-fluorouracil (5-FU) infusion administered through a central venous catheter as an outpatient. Approximately 50% of patients have RAS mutated disease and can be offered non-infusional regimes using oral capecitabine with intravenous oxaliplatin (oral+IV) instead of infusional 5-FU with similar clinical benefit (Cassidy et al, 2008; Ducreux et al, 2011). Patients with BRAF mutations have a significantly poorer prognosis and require a different treatment regimen (BRAF mutant patients have an estimated overall survival of 6-9 months, compared to over 36 months estimated overall survival for RAS and BRAF wild type patients). Therefore, there is a significant impact on the level of information that can be given at a first consultation when the RAS and particularly BRAF mutation status are not known (Yuan et al, 2013). Internal audit data from the West of Scotland also demonstrate that due to the poor prognosis conferred by mutant BRAF

status; these patients are unlikely to reach second line therapy and would be best served by entry into clinical trials or consideration of triplet chemotherapy first line.

As patients are appointed to oncology clinics as soon as possible after the diagnosis of metastatic colorectal cancer, the results of genetic tests are often unavailable at the first oncology appointment. Patients usually expect to discuss their diagnosis, prognosis and management plan at their first appointment, and also expect to start systemic treatment at the earliest opportunity, particularly if they have high symptomatic burden. Without the genetic test results, definitive information about prognosis and the treatment plan cannot be given. To avoid delays in treatment, patients fit for all therapies are often consented for insertion of a peripherally inserted central venous catheter (PICC) and commenced on an IVI chemotherapy regimen while awaiting the RAS test result in case they become eligible for the addition of EGFRi. Even so, patients with RAS wild-type disease will require a further consultation with medical staff in order to consent to the additional EGFRi. Patients with BRAF-mutant disease may have missed the opportunity to participate in clinical trials specific to their disease. Patients with RAS mutant disease are highly likely to remain on IVI chemotherapy once commenced, which is more expensive than the oral+IV alternative as well as being more invasive and inconvenient for patients.

Clinicians suggested that a potential solution to the delays in availability of test results and consequent sub-optimal interventions would be to incorporate KRAS, NRAS and BRAF gene testing into the routine tests undertaken on diagnosis for all patients, including those without metastatic disease. This reflex testing strategy would serve a further purpose of improving the information available to study the RAS and BRAF mutation colorectal cancer patient cohorts as they represent large areas of unmet clinical need. Screening patients in the first line setting offers the opportunity to correlate response to adjuvant chemotherapy, disease free survival and primary tumour site with mutation status in order to advance the standard of care for these patients. Screening of patients at diagnosis for the presence of mutations in RAS and BRAF would also offer the opportunity for increased entry of patients onto clinical trials in the adjuvant setting. Reflex testing of all new colorectal cancer diagnoses would allow the implementation

of a robust system for the collection and processing of deoxyribonucleic acid (DNA) for this patient cohort; thus providing a repository of diagnostic material that can potentially be used for extended testing for patients wishing to enter future clinical trials or for selection of appropriate patient cohorts for future therapies associated with a molecular companion diagnostic marker. However, this extension to testing would increase costs for the nationally-funded molecular pathology service so an economic evaluation of the change in downstream treatment costs and health outcomes was requested.

The aim of economic evaluation is to consider the costs and consequences of a course of action in order to determine whether this use of resources is better than the next available alternative (Drummond et al, 2015). The objective of this study was to consider whether the strategy of reflex testing all CRC patients for RAS and BRAF status on diagnosis would be cost-effective compared to testing patients on request.

9.2 Methods

This case study comprises a clinical value assessment and an economic value assessment. The initial clinical value assessment was based on a clinical audit of the pathways for CRC molecular pathology testing. Further evidence generation took place through the activities of a multi-disciplinary panel and mining of routine data sources. As the panel were satisfied that the extension in testing would only lead to an improvement in health outcomes, the economic value assessment took the form of a cost-minimisation analysis.

The initial step in the process was to form a multi-disciplinary project team including molecular pathologists, clinician and health economists. An employee of Merck, the manufacturer of cetuximab was also part of the project team as the company collect a substantial body of evidence on CRC and are knowledgeable about molecular pathology practice on an international basis. This multi-disciplinary team was important to ensure that all relevant considerations were included. The current and proposed treatment pathways for newly diagnosed patients were mapped in outline using clinical guidelines from the National Institute for Health and Care Excellence (NICE, 2014) and the Scottish Intercollegiate Guideline Network (SIGN, 2018) and input from

clinicians. These pathways were set out as a decision tree with a 26 week time horizon as this is in line with first-line treatment duration. This time-horizon was felt to be sufficient to capture any differences in costs. A UK NHS perspective was adopted. Clinical outcomes were not modelled as in the judgement of the project team clinical outcomes would only be improved by the change from the current to the proposed pathways. This is because outcomes from oral and infusional chemotherapy regimens are equivalent (Cassidy et al, 2008; Ducreux et al, 2011) and outcomes would improve due to reduction in adverse events and reduction in quality of life associated with the insertion of a PICC line.

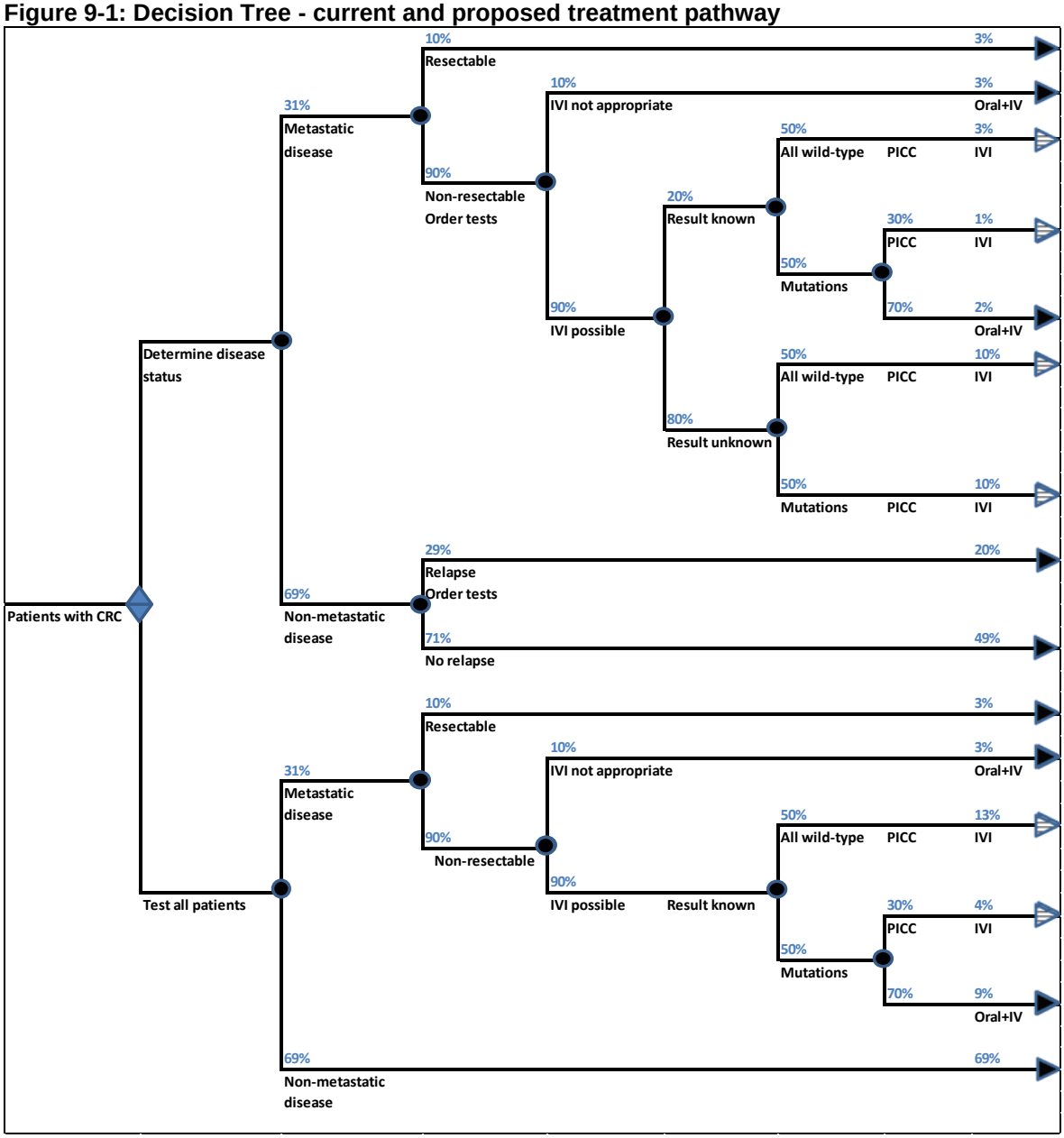
Figure 9-1 shows both the current pathway whereby patients are tested for NRAS/KRAS/BRAF at confirmation of metastatic disease and the proposed pathway where all patients are tested at diagnosis. The first split of the tree indicates the decision whether to follow the current or the proposed pathway. The second split in the current pathway shows the division between metastatic and non-metastatic disease. For patients with metastatic disease the next split divides those who have resectable (operable) disease from other metastatic patients. For some patients with unresectable disease IVI chemotherapy would be inappropriate, for example due to patient fitness, and this is shown as the third split in the decision tree. For the remainder of patients with unresectable disease, some have a RAS/BRAF result known at their first oncology appointment. If the result is known, PICC lines are inserted appropriately and chemotherapy started according to NRAS/KRAS status. Where the result is unknown, PICC lines are inserted and IVI chemotherapy commenced (in case EGFRi can be added once the result is known). For patients with non-metastatic disease at diagnosis, the model only includes the cost of later NRAS/KRAS testing which is done for patients who later relapse. Where results are not known at the initial clinic visit a further clinic visit is required to inform the patient of test results and implications. This is not represented in Figure 9-1 to retain clarity of presentation.

In the proposed pathway (the lower branch of the decision tree) all patients are tested for NRAS/KRAS/BRAF status at diagnosis of CRC (reflex testing). This allows a simplification of the metastatic, unresectable disease branch as all

results are known. In the proposed pathway, all patients with no mutations (all wild-type) will have PICC lines inserted and receive IVI chemotherapy. Some patients with mutations will continue to receive IVI chemotherapy where the clinician recommends this.

Data to populate the decision trees was taken from national and regional cancer registries, a comprehensive Health Technology Appraisal on treatment with cetuximab (NICE, 2018a), SMC Advice on cetuximab (Scottish Medicine Consortium, 2018b) , UK NHS reference costs (NHS, 2018) and a micro-costing of current and proposed laboratory costs undertaken by the project team. Some assumptions were made by the project team based upon their clinical knowledge (proportion of patients who would be clinically unable to receive IVI chemotherapy, the proportion of patients whose cancer was metastatic but wholly resectable (i.e. both primary tumour and metastatic sites operable) at diagnosis, and the proportion of patients whom clinicians would choose to treat with an IVI regimen regardless of whether they would later qualify for treatment with cetuximab). Inputs to the model and data sources are set out in Table 9-1. PICC and Hickmann line insertion and maintenance incurs costs which include; out-patient appointment for siting of line, imaging to ensure placement is correct and district nurse support in the form of twice weekly visits for flushing and maintenance of the line and port. PICC and Hickmann lines are also associated with adverse side effects such as bleeding, clots, infection risk and slippage of line requiring re-siting. The costs included for these lines are based on the costs of PICC line insertion, removal and maintenance. Costs of adverse events are not included in the model. Costs are not discounted given the short time horizon of the economic evaluation.

Figure 9-1: Decision Tree - current and proposed treatment pathway



PICC – Peripherally inserted central catheter, IVI – intravenous infusional chemotherapy, Oral+IV – Oral and intravenous chemotherapy, CRC – colorectal cancer. Diamond represents a decision node. Circle represents a probability node. Percentages represent the base case estimates (see Table 9-1). Solid arrow represents branches where treatment pathways are identical between the current and proposed strategies, striped arrows represent changed treatment pathways

One-way sensitivity analysis was undertaken whereby certain input parameters were varied in turn over a given proportion (whilst all other inputs to the model were held constant) to determine the impact of an over or under-estimation on the base-case results. This form of sensitivity analysis was undertaken as this allowed decision makers to assess the individual impact of each of the input parameters. One way sensitivity analysis was undertaken for 1) the proportion of patients metastatic at diagnosis (range 20% -55% taken from NICE Technology Appraisal 242) (NICE, 2018a) (2) proposed test costs (range from base case £120

calculated in micro costing exercise to £200 which is the level reimbursed previously by Merck to laboratories carrying out these tests and considered to include an element of surplus over cost) and 3) the proportion of patients whose disease was not metastatic at diagnosis but who subsequently relapse (range 29% reduced to zero).

Where no suitable range could be identified from evidence sources the range was varied to determine the highest level at which the cost savings would be reduced to zero. This threshold analysis was undertaken for the proportion of patients who are NRAS/KRAS/BRAF wild type, the proportion of patients with resectable metastatic disease, the proportion of patients for whom IVI chemotherapy would be inappropriate, the proportion of patients prescribed IVI chemotherapy regardless of NRAS/KRAS/BRAF status and the proportion of results currently known at the multi-disciplinary team meeting.

Some input variables were not varied as the project team was confident that the value was appropriate and supported by good quality evidence. These variables were incidence of CRC in the West of Scotland which is supported by registry information and the cost of FOLFOX and XELOX (IVI and oral+IV chemotherapy treatments) which was obtained from a comprehensive Technology Assessment report (NICE, 2018a). The costs relating to PICC lines were also not subject to sensitivity analysis as they are believed to be under-estimated by the project team. This is primarily because maintenance of a PICC line used in the West of Scotland in CRC requires two visits per week by a district nurse which would cost more than the £63 allowed in this analysis. Moreover, no costs are included for the adverse events associated with PICC lines, such as blockage and infection.

Table 9-1: Inputs to the model and data sources

Epidemiology	
Proportion of patients metastatic at diagnosis [a]	0.31
Proportion of metastatic patients with KRAS/NRAS wild-type (Scottish Medicine Consortium, 2018b)	0.50
Proportion of patients for whom IV treatment is not appropriate [b]	0.10
Proportion of patients with mutations who receive IVI chemotherapy for other clinical reasons [b]	0.30
Proportion of non-metastatic patients at diagnosis who will relapse [a]	0.29
Proportion of metastatic patients with resectable primary and metastases [b]	0.10
Proportion where RAS/KRAS/BRAF status known at first clinic visit [c]	0.20
Incidence of colorectal cancer in the West of Scotland 2015	1,606
Costs	£
Cost of current KRAS/NRAS/BRAF test [d]	120
Cost of proposed KRAS/NRAS/BRAF test [d]	120
Cost of oral+IV chemotherapy (26 weeks) (Scottish Medicine Consortium, 2018b)	5,832
Cost of IVI chemotherapy (26 weeks) (Scottish Medicine Consortium, 2018b)	9,893
PICC line insertion (NHS, 2018)	377
PICC line removal (NHS, 2018)	176
PICC line maintenance (NHS, 2018)	63
Total PICC line cost [e]	2,191
Clinic visits (NHS, 2018)	197

Number in brackets is the reference to the source of the data where a reference is included. Letters refer to the notes below:
[a]Data provided by David Creed, co-author (unpublished)
[b]Expert opinion from cross-disciplinary project team
[c]Data from internal audit performed by Janet Graham and Yun Yi Tan (co-authors)
[d]Data from micro-costing analysis undertaken by Gillian McGaffin (co-author) at Glasgow Molecular Pathology Laboratory (2018)
[e]Calculated as insertion+removal+(26xmaintenance)
IVI – intra-venous infusional chemotherapy regimen, oral+IV – oral and intravenous chemotherapy regimen, PICC – peripherally inserted central catheter

9.3 Results

The base-case analysis indicated that the process change would save £397 per patient which equates to £637,332 per annum in the West of Scotland. This saving results primarily because approximately 7% of patients (n=112) were predicted to avoid PICC line insertion and IVI chemotherapy which more than outweighs the additional cost of testing for those patients with non-metastatic disease at diagnosis and who do not relapse. Those avoiding PICC line insertion and IVI therapy can be seen from Figure 9-1 as in the current strategy the striped arrowheads showing IVI treatment total 24% (3%+1%+10%+10%), whereas under the proposed strategy IVI treatment totals only 17% (4%+13%). Additional testing costs are incurred for 49% patients being the difference between testing 100% patients under the proposed strategy compared to 51% (31% metastatic plus 29% relapsed of 69% non-metastatic) under the current strategy. Testing costs are anticipated to reduce under the new testing strategy as a result of increased volumes and this test cost has been included as both current and proposed test cost in order to ensure estimates of cost savings are conservative.

Sensitivity analysis showed that the overall cost saving was not highly sensitive to any individual parameter when varied within ranges judged feasible by the project team. The threshold analysis showed that the assumptions made about the proportions of patients falling in each category could vary considerably before the overall cost saving was reduced to zero. This is because the cost saving from diverting an individual patient from IV treatment and PICC insertion is high compared to test costs so that only a small number of patients need to be diverted in order for the proposed change to deliver cost savings. Table 9-2 summarises the base case results and the results of sensitivity and threshold analyses.

Table 9-2: Base-case results, sensitivity and threshold analysis

Base case	£
Cost saving per patient	397
Cost saving for the West of Scotland (annual incidence n=1,606)	637,332
Additional costs of testing (49% of 1,606 patients at £120 per test)	94,433
Cost savings from PICC lines (7% of 1,606 patients at £2,191)	246,312
Cost savings from treatment with oral+IV rather than IVI chemotherapy (7% of 1,606 patients at £4,061 (£9,893-£5,832))	456,538
Cost savings from additional clinic visit (7% of 1,606 patients at £197)	22,147

Sensitivity analysis	**Cost saving in £**
Proportion of patients metastatic at diagnosis [range 20% -55%] (Scottish Medicine Consortium, 2018b)	362,629 – 1,236,684
Proposed test costs [increased from £120 to £200]	570,407
Proportion of patients whose disease was not metastatic at diagnosis but who subsequently relapse [reduced from 29% to 0%]	549,769

Threshold analysis	**Base case**	**To reduce cost savings to zero**
Proportion of patients who are NRAS/KRAS/BRAF wild type	50%	93%
Proportion of patients with resectable metastatic disease	10%	82%
Proportion of patients for whom IVI chemotherapy would be inappropriate	10%	87%
Proportion of patients prescribed IVI chemotherapy regardless of NRAS/KRAS/BRAF status	30%	90%
Proportion of results currently known at the multi-disciplinary team meeting	20%	88%

9.4 Discussion

A strength of this study is that it demonstrates the use of economic evaluation to assess a change of process at a regional level. This change of process is likely to be cost saving for the region and suggests that this approach may offer a means of delivering process change, efficiencies and perhaps disinvestment on a wider scale within healthcare services. Key aspects of the study in this regard were the local focus, the buy-in of a cross-disciplinary expert project group representing the different service areas affected and the presentation of both qualitative and quantitative information to the decision makers.

Previous adaptations to molecular pathology testing in Scotland have not been supported by health economic analysis. This analysis was made possible through the involvement of Glasgow Molecular Pathology Node (GMP Node), a collaboration between the University of Glasgow, NHS Greater Glasgow and Clyde and commercial partners, funded by the Medical Research Council and the Engineering and Physical Sciences Research Council. This study demonstrates the value of such translational research bodies in facilitating the introduction of molecular pathology tests into clinical practice. The GMP Node provided a forum for clinicians and pathologists to collaborate and funded the health economics aspect of the study.

Further strengths of the study are that it was relatively quick and resource light, as a result of the suitability of a costs-only analysis, the simplicity of the model and the approach to evidence gathering. A costs-only analysis was possible as outcomes were only likely to be improved by the availability of more information or earlier information. The model was simple as it considered only costs and had a relatively short time horizon. Evidence was based on easily available national resources and local experience informed by the expert project team. One-way sensitivity analysis was appropriate as the cost saving was relatively large and not sensitive to any individual parameter. It also allowed decision makers to assess the importance of individual parameters. Had the result been more sensitive then multi-way sensitivity analysis may have been appropriate. Probabilistic sensitivity analysis was not appropriate given the number of assumptions made by the project team as this may have given an impression of false precision.

A limitation of the study was the extent to which the modelling relied upon assumptions made by the expert project team. This study will be followed by a practical pilot project which will assess the validity of the assumptions. Given the fast-moving environment of molecular pathology testing this change in process is likely to happen in any event if trials of PDL1+/- CTLA4 inhibitors become the standard first-line metastatic treatment for therapy for patients with micro-satellite instability high metastatic cancers (National Cancer Institute, 2018). The use of these same agents in the adjuvant setting is beginning to be investigated in clinical trials such as POLEM (Clinical Trials Register, 2017).

The decision to be informed in this study was whether the extension of the pathology testing for CRC in the West of Scotland should be funded. This study found that an extension of testing for RAS/BRAF to all patients diagnosed with CRC in the West of Scotland is likely to deliver cost savings as reductions in downstream treatment costs would more than outweigh the additional cost of testing. Although the same finding may be true in other locations the study's findings are not directly generalizable as they are dependent upon existing local treatment pathways, the organisation of local cancer services and capacity and organisation of molecular pathology services. The savings are delivered because cancer services in the region are dispersed across a number of sites with key clinicians providing services in more than one location. The result of this is that organisational change is more complex and other ways of achieving the cost savings are not possible. By way of contrast, molecular pathology services are centralised in one laboratory with a high throughput so that economies of scale can be achieved.

The immediate implication of this study for policy-makers is that, dependent on local context, downstream treatment efficiencies may be realised through changes to molecular pathology testing arrangements. More generally, economic evaluation coupled with a cross-disciplinary project group may offer a method of increasing efficiency of treatment pathways. This example was relatively straightforward as outcomes could be assumed to be improved and the process change could be justified on cost saving alone. Other situations are likely to be more complex. In this study, an increase in spending is required in one area of

the treatment pathway in order to deliver savings elsewhere. In order for this approach to be useful in delivering treatment pathway efficiencies policy makers will need to consider the incentives of all budget holders across the treatment pathway and other stakeholders. In Scotland, incentives are more straightforward than in many other contexts, although, even in this nationally funded system, test costs are borne nationally whereas savings will be realised by the regional health boards who meet the costs of cancer treatment. Other aspects which may be of importance in certain local contexts are capital expenditure required and any capacity constraints. In this case, no capital expenditure was required and capacity could be increased in Glasgow Molecular Pathology Laboratory; these aspects were costed as part of the micro-costing analysis.

Next steps for the developers in this case include further clinical audit to ascertain the veracity of the assumptions made in this analysis. This audit is likely to extend outside of the West of Scotland region to determine whether the finding is applicable to other areas of Scotland. Given the fast-moving nature of the molecular pathology testing environment, MPEP will be required to review testing requirements for CRC again in the near future. It will be important for any review to be supported by an appropriate level of health economic analysis to ensure full account of the implications of the changes is being taken.

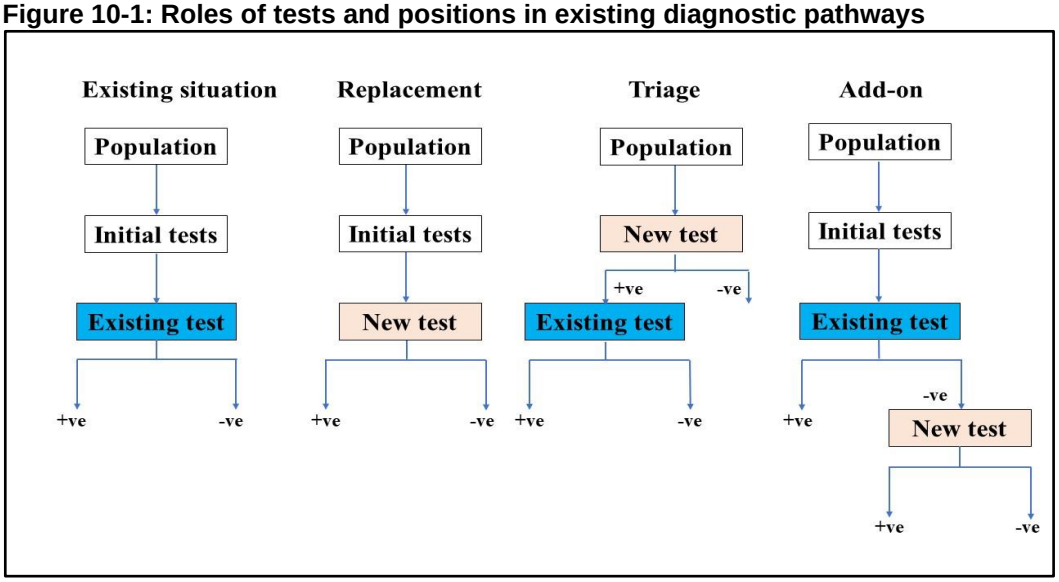
10 Case study 4 - Assessment of triage tests when existing test capacity is constrained: application to an imaging test to aid diagnosis of malignant melanoma

10.1 Introduction

An adapted version of this chapter, entitled 'Evaluation of triage tests when existing test capacity is constrained: application to rapid diagnostic testing in CoVID-19' co-authored by Neil Hawkins, is in preparation.

Where development is technology-driven (i.e. the technology has been developed without a clear setting or position in the clinical pathway identified), a first step in assessment is the consider the optimum positioning of the technology. Diagnostic tests sometimes replace existing tests, but they also may be added to the diagnostic pathway or may act as a triage test where the result of the new test determines which patients undergo an existing test. This is illustrated in Figure 10-1 below.

Figure 10-1: Roles of tests and positions in existing diagnostic pathways



+ve – test positive, -ve – test negative. Reproduced from Bossuyt et al (2006) with permission.

The introduction of a triage test does not aim to improve the diagnostic accuracy of the current pathway. Rather, it reduces the use of existing tests that may be more invasive, cumbersome or expensive than the existing test or for which the patient may need to wait (Bossuyt et al, 2006). An example of a

triage test is a blood test measuring levels of D-dimer in patients with pulmonary embolism (Bossuyt et al, 2006; Page, 2006). The 'reference' diagnostic test is computed tomography (CT), which is highly accurate. However, CT is expensive, requires skilled staff to undertake it and capacity is constrained, so patients may need to wait before they are tested. Although D-dimer has low specificity (around 50%), it has high sensitivity and negative predictive value (above 99%) (Page, 2006). This means that although D-dimer does not pick up all patients who are disease negative, when there is a negative result, it is highly likely that the patient is disease negative. As the test is cheap and there is no waiting time, D-dimer has potential as a triage test to rule out pulmonary embolism.

Tests may be classified as either "rule-in" tests, intended to confirm that a patient has a specific disease or as 'rule-out' tests, intended to confirm that a patient does not have a disease. "Rule-in" tests require high specificity and "rule out" tests, high sensitivity. This relationship is known as the SPIN and SNOUT rule (SPecific test when positive rules IN the disease - SPIN and Sensitive test when Negative rules OUT the disease' – SNOUT) (Baeyens et al, 2019). The extent to which a 'rule-in' or 'rule-out' test is preferred will depend on the prior probability of disease and the consequences, both to the patient and in terms of healthcare resource use, of false positives and false negatives. In this chapter, the use of 'rule-in' and 'rule-out' test in triage testing is explored in contexts where the existing 'reference' test capacity is both constrained and unconstrained. The chapter sets out to demonstrate that, if facilities for reference testing are restricted, there may be a role for a triage test to enrich the prevalence of disease in the tested population and improve the efficiency of the reference testing process. A simple modelling approach is presented which can be used to explore the use of a triage test prior to a reference test. Two situations are contrasted where the availability of the reference test is capacity constrained or non-constrained. It is shown that the relevant metrics of test value, and the value of a test with given characteristics, vary depending on whether existing reference test capacity is constrained or not. Generic models are developed then applied in a case study using test data for a diagnostic test for malignant melanoma. The models presented in this chapter are not intended to provide definitive estimates of the clinical value and cost-effectiveness of a test, this typically requires the careful identification and synthesis of evidence

and the development of detailed decision-analytic or cost-effectiveness models which requires both time and significant resource. Such models are often highly context specific. Rather, the models presented here are simple models that can be readily used to provide an indication of the potential value of a test whilst it is under development or during some form of expedited review. Such models are likely to be useful during the development of a test and provide a guide as to whether further investment in the development of a test is warranted and what studies are required to provide sufficient evidence to support the uptake and commercialisation of a test.

The context for the case study is malignant melanoma (MM). Here, the diagnostic pathway in the UK involves screening by primary care practitioners and onward referral to specialist dermatologists. Due to the number of referrals from primary care, meeting the two-week targets set for specialist dermatologist review in England is challenging. In some areas these targets are not being met. Even where they are being met, the level of demand on specialists' time may result in patients with other skin conditions being forced to wait. As many of the patients referred with suspected MM will ultimately prove to be negative, there is a high opportunity cost in terms of health lost whilst patients with other conditions are waiting for specialist appointments.

A commercial company developing an imaging test using machine learning to identify MM approached Neil Hawkins to undertake health economic assessment. The company's initial aim had been to offer a test in primary care which would safely prevent onward referral of a proportion of the suspicious moles. They had carried out an initial clinical study involving patients who had been referred to specialist dermatologists.

10.2 Methods

A decision analytic model was developed that predicted true and false positive and negative rates for a triage test (T) with given sensitivity (SENS), specificity (SPEC) and prevalence. In this case, sensitivity and specificity were defined strictly in terms of the probability that the triage test predicts positive and negative test results for a patient receiving the existing or 'reference' test that is used to determine the future treatment of a patient. In this notation, $P[R^+]$

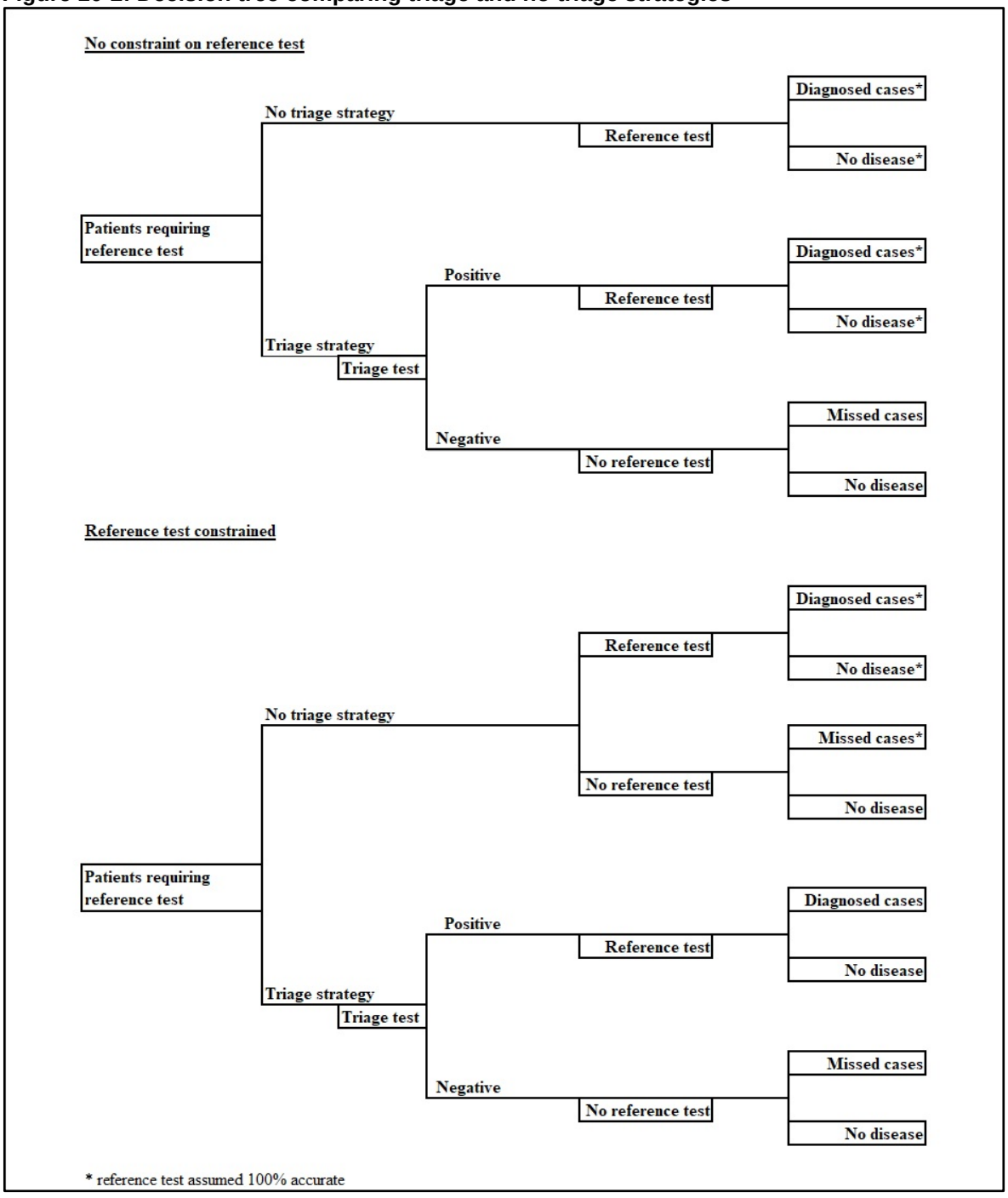
(probability of testing positive in the reference test) is effectively the prevalence as it is assumed that the reference test is 100% accurate. It is assumed that a patient testing positives or negatives with the reference test is disease positive or disease negative respectively.

$$SENS = P[T^+|R^+]$$

$$SPEC = P[T^-|R^-]$$

The model is shown in Figure 10-2.

Figure 10-2: Decision tree comparing triage and no triage strategies



Two ‘use cases’ are considered. One where the reference test capacity is constrained and currently fully utilised and one where the reference test capacity can be varied in the short term according to demand.

10.2.1 Reference test capacity unconstrained use case

Where the reference capacity is unconstrained, the reduction in the number of reference tests undertaken and the number of false negative results arising from the use of the triage tests can be estimated. The reduction in the proportion of patients requiring the reference test following the triage test (the probability that a patient tests negative on the triage test) is given by:

$$1 - \left(P[R^+] \times SENS + (1 - P[R^+]) \times (1 - SPEC)\right)$$

In these expressions $P[R^+]$ is the probability of the reference test being positive (equivalent to prevalence in this simple model where the reference test is assumed to be 100% accurate). SENS means sensitivity of the triage test and SPEC, specificity of the triage test. The proportion of missed cases, patients who test negative on the triage test but would have tested positive on the reference test, is given by:

$$P[R^+] \times (1 - SENS)$$

The reduction in the proportion of patients requiring the reference test and associated cost savings will need to be traded off against the missed cases.

10.2.2 Capacity constrained reference test use case

Where the reference capacity is constrained, rather than comparing the reduction in proportion of patients requiring the reference test and missed cases, the total number of true positive cases that are identified using the triage test is compared with the number identified using a reference test only strategy.

The probability of testing positive at triage testing ($P[T^+]$) is estimated:

$$P[T^+] = P[R^+] \times SENS + (1 - P[R^+]) \times (1 - SPEC)$$

In this expression $P[R^+]$ is the probability of the reference test being positive (equivalent to prevalence in this simple model where the reference test is assumed to be 100% accurate). SENS means sensitivity and SPEC, specificity. This provides an estimate of the proportion of patients for whom reference

testing is indicated following the triage test. The number of patients who can actually be tested will depend on the available capacity.

The relative maximum population size (δ_{triage}) that can be tested using the triage testing before the reference test capacity is exhausted is then estimated:

$$\delta_{triage} = \frac{1}{P[T^+]}$$

The probability of testing positive at the reference test following triage testing is then calculated:

$$P[R^+|T^+] = \frac{P[R^+] \times SENS}{P[R^+] \times SENS + (1 - P[R^+]) \times (1 - SPEC)}$$

This assumes that there are sufficient untested patients under the status quo to exhaust reference test capacity following triage testing. To compare the *status quo* and triage test scenarios the number of identified cases and missed cases from the expanded populations is estimated. For the reference test only strategy, the proportion of identified cases from the expanded population is given by:

$$\frac{P[R^+]}{\delta_{triage}}$$

For the triage test strategy, the proportion of identified cases from expanded population is given by:

$$\frac{P[R^+|T^+]}{\delta_{triage}}$$

The difference between these proportions represents patients who would have tested positive under the reference test, but were unable to receive the test due to the capacity constraint. It is assumed that $P[R^+]$ is the same for patients who were able to, and not able to, access the reference test. Finally, the

increase in the probability of identifying a case from the expanded population that arises from the use of the triage test can be estimated. This increased probability of detecting a case can be traded off against the cost of the triage test.

$$\frac{P[R^+|T^+]}{\delta_{triage}} - \frac{P[R^+]}{\delta_{triage}}$$

10.3 Case study

10.3.1 Background

10.3.1.1 Epidemiology

In 2016, malignant melanoma was the fifth most common cancer for males and females in Scotland (ISD Scotland, 2018a). The incidence rate for males was 28.9 per 100,000 years at risk and for females 24.6. The corresponding mortality rates were 4.6 for males and 2.1 for females. Incidence rates for males increased by just under 30% over the last decade. For females, there was no evidence of an increasing trend over the past decade but substantial increases had occurred previously (ISD Scotland, 2018a) including a 7-fold increase between 1976 and 2009 (NICE, 2015). MM is most common in people over 50 years of age but 20% of cases occur in young adults (NICE, 2015). 80% of cases are thought to be linked to ultra-violet exposure from the sun and sunbed use (NICE, 2015). Incidence of MM is lower in lower socio-economic groups (NICE, 2015). Cancer survival for MM is relatively good as most cancers present at an early stage (ISD Scotland 2018a). Prognosis is poor for cancers presenting at a late stage although advances in treatment have been made in recent years (SIGN, 2017).

10.3.1.2 Clinical pathway

Early detection of melanoma is important as early treatment is usually curative (SIGN, 2017). Suspicious lesions are generally noticed either by the patient or a family member or by a clinician in the course of another consultation (SIGN, 2017). The current diagnostic pathway in Scotland, requires the clinician in

primary care to urgently refer the patient if any of the major features of the 7-point checklist lesion system (Table 10-1) or any of the features in the ABCDE system (Table 10-2) are present. The presence of the minor features from the 7-point checklist should increase suspicion (SIGN, 2017). Diagnosis of melanoma is difficult (SIGN, 2017). There is evidence that the accuracy of diagnosis varies according to a clinician's experience and that there is considerable variation in the sensitivity of the diagnosis (SIGN, 2017). In secondary care, assessment using dermoscopy by trained healthcare professionals is recommended but the precise pathway may vary by healthcare setting (SIGN, 2017). Positive dermoscopy leads to biopsy and pathological analysis (SIGN, 2017).

Table 10-1: The 7-point checklist lesion system (SIGN, 2017)

Major features	Minor features
Change in size of lesion	Inflammation
Irregular pigmentation	Itch/altered sensation
Irregular border	Lesion larger than others
	Oozing/crusting of lesion

Table 10-2: The ABCDE lesion system (SIGN, 2017)

A	Geometrical **A**symmetry in two axes
B	Irregular **B**order
C	At least two different **C**olours in lesion
D	Maximum **D**iameter >6 mm
E	**E**volution/change in lesion

In the UK, various targets are in place to try and improve health outcomes for patients diagnosed with cancer. In England and Wales, there is a two-week target for referrals from primary care to first appointment in secondary care. In Scotland, there is a 62-day target from referral to first cancer treatment and a 31-day target from the decision to treat to treatment starting (ISD Scotland, 2018b). In the quarter to 30 June 2018, 97.7% of patients with melanoma started treatment within the 62-day target and 95% of patients were treated within 31 days of the decision to treat (ISD Scotland, 2018b).

Referrals cause a heavy burden on specialist dermatology services and a high proportion of referrals for suspicious lesions are benign. An overview of 52 audit studies of skin cancer referrals under the 2-week rule found that melanoma and squamous cell carcinoma accounted for only 10% to 12% of referrals, with the remainder being benign lesions (CCG Cancer Assessment, 2017/18). In England,

there has been exploration of the potential of teledermatology (Eastham and O'Shea, undated; PCC, 2013) to combat this. Teledermatology involves the use of imaging technology in primary care to allow the triage of suspected cases of skin cancer including melanoma. One study suggested that up to 50% of referrals could have been avoided if teledermatology had been used in primary care prior to referral (Cox, 2004).

10.3.1.3 Technology and initial clinical study

The technology assessed in this case study is the melanoma test (name changed for confidentiality reasons). It is a photographic image analysis system using iPad technology (iPad Air2 device) and proprietary signal processing (using the melanoma test app) (melanoma test clinical study report 004, unpublished, hereafter 'CSR4'). The software creates numerical values for five defined characteristics of the image, some of which are not visible to the naked eye (CSR4). Library images of benign and malignant pigmented lesions were used by the developers to establish thresholds within an algorithm which were then tested in a hospital-based study in three hospitals (Wishaw General Hospital, University Hospital, Hairmyres and University Hospital, Monklands, all in Lanarkshire, Scotland)(registered clinical trial protocol and CSR4). 1,200 patients were recruited over a 12-month period from April 2015 from the population of patients referred to specialist dermatologists from primary care with suspicious lesions. Data from 400 patients was used to calibrate the software and data from the remaining 800 was used for validation. The classification by the melanoma test was compared to clinical diagnosis by inspection and biopsy result, if one was performed (CSR4). A population, intervention, comparator, outcome (PICO) summary for CSR4 is set out in Table 10-3.

Table 10-3: PICO summary for clinical study report 4 (unpublished)

Population	Patients referred to specialist dermatologists with suspicious lesions
Intervention	Test with melanoma test, +ve result = referred, -ve result = not referred
Comparator	All patients referred
Outcome	Proportion of patients correctly referred Referrals avoided

PICO – population, intervention, comparator, outcome

Four possible categories of diagnosis are of interest for the clinician. These are (from least to most serious: benign, dysplasia, other malignancies and

melanoma. The developers investigated four alternative thresholds (termed 'discriminant functions') which are set out in Table 10-4. D1 is the narrowest discriminant function and is set as if only MM is of interest to clinicians. Each subsequent function will capture more of the other results of interest (melanoma, other malignancies and dysplasias) but will also include a greater proportion of false positives and will not reduce the number of referrals to the same extent. D1 and D2 were selected as the discriminant functions of interest in order to compare the relative performance of D1 (rule-in test with high specificity relative to sensitivity) and D2 (rule out test with high sensitivity relative to specificity) in conditions with and without constraint on specialist referral.

Table 10-4: Discriminant functions specified by developers

Discriminant function	Dysplasia	Other malignancies	Melanoma
D1			At least one measurement beyond the range for melanoma
D2		At least one measurement beyond the range for other malignancies	At least one measurement beyond the range for melanoma
D3	Excluding 90% of dysplasias	At least one measurement beyond the range for other malignancies	At least one measurement beyond the range for melanoma
D4	Excluding all dysplasias	At least one measurement beyond the range for other malignancies	At least one measurement beyond the range for melanoma

Shading indicates the discriminant function applied in this chapter.

Table 10-5: Test performance for discriminant functions D1 to D4

Threshold	Sensitivity	Specificity
D1 - at least one measurement beyond the range for melanoma	46%	63%
D2 - at least one measurement beyond the range for melanoma and for other malignancies	82%	23%
D3 - at least one measurement beyond the range for melanoma and for other malignancies and excluding 90% of dysplasia	92%	18%
D4 - at least one measurement beyond the range for melanoma and for other malignancies and excluding all dysplasia	100%	8%

Shading indicates the discriminant functions of interest in this chapter.

10.3.2 Method

Decision tree models were developed which compared the current (refer all) and alternative test and refer strategies for both non-constrained and constrained resource scenarios (identical to the generic case in Figure 10-2). In the proposed pathway (the higher branch of the decision tree), all patients are tested using the melanoma test. Those with a positive result are referred to a specialist

dermatologist. Those who do not meet the threshold (negative result) are not referred. In both positive and negative test result arms there is the possibility of the test result being correct or incorrect resulting in missed cases of dysplasia, other malignancies and melanoma. The only difference between the constrained and non-constrained decision trees is the constraint on specialist dermatologist appointments. This means that a proportion of patients referred in the 'no triage test' arm will not see the specialist dermatologist so no diagnosis will be made.

The time horizon for the decision tree covered the diagnostic process up to referral. Inputs to the model include the sensitivity and specificity associated with the D1 and D2 discriminant functions (see Table 10-5) as well as the prevalence of all the conditions of interest (23.5% from CSR4). For the purposes of this analysis, test accuracy is assessed as the ability to distinguish all the conditions of interest. So 82% sensitivity (see Table 10-5) means that at a threshold set to identify all suspicious lesions with at least one measurement beyond the range for melanoma or other malignancies the test will only identify 82% of lesions with MM, other malignancies or dysplasia.

10.3.3 Results

Table 10-6 sets out the results of a reference test only strategy with a reference plus triage test strategy including melanoma test as both rule in and rule-out tests in the event that specialist dermatologist appointments are not constrained.

Table 10-6: Results with dermatologist appointments non-constrained

Strategy/Parameters Population = 1,000	Specialist dermatologist appointment only	Melanoma test as triage test plus specialist dermatologist appointment	
		Rule out test (Sens 82%, spec 23%)	*Rule in test (Sens 46%, spec 63%)*
Positive at triage	-	78.2%	39.1%
False negative rate at triage	-	4.2%	12.7%
Reduction in reference testing	-	21.8%	60.9%
Percentage of reference tests positive	23.5%	24.6%	27.6%
Number of reference tests required	1000	782	391
Missed cases	0	42	127

Sens – sensitivity, spec - specificity

Table 10-7 sets out the results assuming appointments are constrained. For the reference capacity constrained case, the relative expansion of the population and the number of cases identified is greater with the higher specificity, 'rule-in' test. Although, the number of false negatives at triage is also greater with the 'rule-in' test due to the lower sensitivity, this does not equate to missed cases. The total number of cases identified is still higher with the rule-in test due to the expansion of the tested population. For the unconstrained reference capacity case (see Table 10-6), the higher specificity 'rule-in' test leads to the greatest reduction in the number of reference tests required, however the 'rule-in' test also leads to the greatest number of missed cases. The higher sensitivity 'rule-out' test minimises the number of missed cases but is less efficient in reducing the number of reference tests.

Table 10-7: Results with dermatologist appointments constrained

Strategy Parameters Capacity constraint = 1,000	Specialist dermatologist appointment only	Melanoma test as triage test plus specialist dermatologist appointment	
		Rule-out test (Sens 82%, spec 23%)	*Rule in test (Sens 46%, spec 63%)*
Probability test positive at triage	-	78.2%	39.1%
False negative rate at triage		4.2%	12.7%
Percentage of reference tests positive	23.5%	24.6%	27.6%
Maximum relative expansion of coverage[1]	-	1.28	2.56
Maximum number of patients triage-able before reference test capacity (1,000) exhausted	-	1,279	2,557
Number of positive reference tests	235	246	276
Incremental increase in cases found due following triage	-	11	41
False negatives at triage	-	54	324

Sens – sensitivity, spec - specificity

The value of the melanoma test technology as a rule in or rule out triage test can be estimated assuming specialist dermatology resource was not constrained (see Table 10-8). The cost of £137 for a specialist dermatology referral was taken from public sources relevant for Scotland (Curtis and Burns, 2017). It was not discounted given the short time horizon of the economic evaluation.

Table 10-8: Results using the melanoma test at D1 and D2 thresholds

Strategy	Referrals avoided (all patients who test negative at triage)	Missed cases - all	Value of test
Refer all	0%	0%	N/A
Triage using D1 – rule in - at least one measurement beyond the range for melanoma	61%	13%	£84
Triage using D2 – rule out - at least one measurement beyond the range for melanoma and for other malignancies	22%	4%	£30

When resource is constrained the value of the triage test is in increasing the number of cases diagnosed rather than in reducing the number of specialist dermatologist appointments required. An economic value assessment would need to compare the downstream costs and outcomes resulting from each strategy in order to determine whether a triaging strategy was potentially cost-effective.

10.4 Discussion

Where the reference test capacity is constrained, the assessment of the opportunity cost of a triage strategy depends on the cost of the triage testing and an assessment of the value of the additional cases identified. The number of additional cases identified is maximised by maximising specificity. Arguably sensitivity is less important as false negatives at triage testing do not directly represent missed cases. As long as there is a sufficient pool of individuals who would not be tested under the reference test only regime, the total number of cases identified will be greater under the triage strategy. However, there still may be a 'cost' associated will false negatives at triage testing if these lead to a change in behaviour in those who test negative at triage.

Where the reference test capacity is unconstrained, the assessment of the opportunity cost of a triage strategy depends on the cost of the triage testing, the costs resulting from a reduction in reference testing, and the 'cost' of any false negatives at triage testing in terms of worse outcomes. The choice of test

will depend on trade-off between cost-savings and missed cases due to false negatives at triage.

When the availability of the gold-standard diagnostic test is constrained, tests added to the diagnostic pathway in a triage position do not necessarily need high levels of accuracy in order to increase the number of cases diagnosed. This has implications across a wide range of disease areas. The model presented in this chapter provides a simple way of assessing whether a particular test may have potential to increase diagnoses. This chapter has shown, using a simple model, that the value of a triage test, in terms of additional cases diagnosed, depends upon whether the availability of the gold-standard (or reference) test is constrained. Where resource is not constrained, a triage test will result in missed cases and an overall reduction in diagnoses unless its accuracy is high. However, where the availability of the reference test is constrained, there is potential for tests with relatively low accuracy to improve levels of diagnosis and reduce missed cases. The levels of test performance required depend upon the extent to which the capacity of the reference test is constrained. Levels of test performance can be lower and still deliver benefits as the reference test numbers available reduce as a proportion of the population to be tested. This finding has relevance across a wide range of diseases and settings. For example, in tuberculosis approximately 3.6 million cases of active disease go undiagnosed annually, partly due to limited access to confirmatory molecular tests (Nathavitharana et al, 2019). A further example is in colorectal cancer. In the UK, demand for colonoscopy is forecast to increase 10-15% year on year resulting in capacity constraints and faecal immunochemical tests have been suggested as a possible triage test in symptomatic patients (Westwood et al, 2017). In relation to the COVID-19 pandemic, testing resources have been constrained in many jurisdictions. Required performance levels for RDTs have been set sufficiently high (for example, the Medicines and Healthcare Products Regulatory Agency (MHRA) in the UK have set performance standards of minimum 98% sensitivity and specificity for tests used as ‘immunity passports’ and in ‘seroprevalence studies’ (MHRA, undated)) that many test manufacturers have failed to meet them (Adams et al, 2020) and some tests have been returned to manufacturers (Jones, 2020). It has been suggested that there is a need to be creative in testing strategy. (Pettit et al, 2020; Peto, 2020; Rodriguez, 2020).

The results set out in this chapter suggest that there may scope to use tests with lower performance effectively in some testing pathways.

Evaluation of alternative testing strategies using cost-effectiveness analysis typically requires a complex model with parameter estimates for health outcomes and resource use. This resource intensive process takes time and expertise and results may be difficult to generalise as diagnostic and clinical pathways vary across and within jurisdictions. The model presented is not intended to be an alternative to a comprehensive cost-effectiveness analysis. Rather, it is intended as a model that could be used during the development of a test to determine whether further investment is appropriate and to guide design and evidence generation strategy. However, a potential further use of the simple model may be to inform decision-makers responding to the current COVID-19 pandemic, who need to evaluate a large number of tests in a wide range of testing scenarios. The model could help to narrow the range of alternatives to be explored in more detailed modelling

This is the first simple model to demonstrate the benefit of triage tests when availability of reference tests is constrained. The only inputs required for the model are prevalence, sensitivity and specificity, population and number of reference tests available. Assuming the incremental net (accounting for resource use) health benefit for treated positive cases compared with untreated positive cases is greater than the incremental net health benefit for treated negative cases compared with untreated negative cases, maximising the total number of true positive reference tests will maximise net health benefit. In general, treated negative cases will be associated with a negative incremental health benefit as they will be associated with wasted resources and sometimes harm to patients due to unnecessary treatment.

11 Case study 5 - Economic evaluation of blood test to aid diagnosis of prostate cancer in Hong Kong

11.1 Introduction

This chapter is based on an article published in PLOS One in April 2019. The co-authors were Jeremy Teoh, Peter Chiu, Kevin Chan, Chi-Fai Ng (all of whom were involved in the initial clinical study (Chiu et al, 2016)) and Robert Heggie and Neil Hawkins (health economists who were involved in the validation of the model and who read and approved the manuscript). The decision makers for this case study are a team of academic clinicians working in urinary medicine the Chinese University of Hong Kong. The study arose following a clinical trial in which the decision makers examined the relative merits of different diagnostic strategies. The ultimate aim will be to present a case for reimbursement to the public health authority in Hong Kong. However, at this stage, with clinical evidence limited to some retrospective analysis from the previous trial, the academic clinicians are making decisions around whether further research is likely to be worthwhile and if so, what form it should take. The decision makers' underlying objective, for this study is to maximise societal return on investment from future research. The output of this HTA study will constitute a platform for discussion between the academic clinicians and the Hong Kong Public Health Authority about the potential of the test in use in Hong Kong and the evidence required for adoption of the technology.

11.2 Methods

This case study comprises a clinical value assessment and an economic value assessment. As it concerns a commercially-available technology, the iterative process is well advanced. As opposed to simply attempting to articulate a value proposition the study seeks to evaluate that proposition in a given jurisdiction with the use of some technology and context specific evidence.

Methods applied in this case study were a health impact assessment (to assess clinical value) and a cost-consequence analysis (to assess economic value). As

resources were scarce the health impact assessment relied on input from the academic clinician team from Hong Kong to map pathways. Trial data (both published and unpublished) provided input to the cost-consequence model. Cost-consequence analysis was selected as the introduction of the test was expected to deliver cost savings in the avoidance of a proportion of prostate biopsies and their attendant complications. Although there were improved health outcomes from the reduction in biopsy-related complications and the disutility from the biopsy itself the introduction of another imperfect test into the diagnostic pathway carries a risk of false negatives (missed cancers). Key for the decision-maker is how large a cost saving can be delivered at differing levels of acceptance of risk of missed cancer. Although a full cost-utility modelling exercise may be the most accurate way of synthesising these considerations and presenting a case for reimbursement to the Hong Kong Public Health Authority, such a model is expensive to produce. Moreover, simple models can be more effective tools of communication.

A simple generic mathematical model was developed to compare the cost savings to the consequences of missed cases. Cost savings are from the avoidance of downstream diagnostic costs when a new test is introduced to the diagnostic pathway. Missed cases are the result of false negative results of the new diagnostic test. Both aspects vary with test performance.

11.3 Results

Prostate Cancer (PCa) is the second most commonly diagnosed cancer in men worldwide (Mottet et al, 2017). The incidence of PCa in Chinese men is 10 times lower than the rate in men from Western Europe but it has increased rapidly in recent years (Chiu et al, 2016; Chen et al, 2014). Positive biopsy rates are lower in Asian men (15-25%) compared with Western European men (30%) and cancer tends to be diagnosed later (Chiu et al, 2016; Chen et al, 2014). The first steps on the current diagnostic pathway in Hong Kong for suspected PCa are a digital rectal examination (DRE) and the prostate specific antigen (PSA) blood test. In men whose DRE is normal but whose PSA levels are between 4-10 ng/ml the current diagnostic pathway requires a transrectal ultrasound-guided (TRUS) biopsy. Such biopsies are invasive and carry considerable risks of post-procedure complications including fever, acute urinary retention, haematuria and

haemospermia (Chiu et al, 2016). As positive biopsy rates are low, many biopsies are carried out unnecessarily under the current diagnostic set-up (Chiu et al, 2016).

The Prostate Health Index (PHI) is a commercially available blood test manufactured by Beckman Coulter Inc. which has been recently approved by the United States FDA for use in patients with PSA of 4-10 ng/mL and normal DRE. The test uses a combination of different forms of PSA (tPSA, fPSA and [-2] proPSA (p2PSA) and has shown improved ability to predict presence of PCa and clinically significant PCa at biopsy (Le et al, 2010; Loeb and Catalona, 2014). In a recent study reported elsewhere, Chiu et al (2016) found that PHI had a better performance than PSA-based diagnostic models in the Chinese population. The study found that in a population of men aged 55-75 years, if a 10% risk of missing PCa during screening was accepted, 30.3% of biopsies could be avoided using PHI compared to 16% using PSA. If a 20% risk of missing PCa was accepted, 47.5% of biopsies could be avoided compared with 20.8% using PSA-alone. Data from the whole cohort in this study (including those outside the screening age group of 55-75 years) suggested that in the population with a PHI score of under 25, 3.6% would have PCa and 0.5% would have high-grade PCa. For the population with a PHI score between 25 and 35, 7.6% would have PCa and 0.9% high-grade PCa. This suggests that it may be possible to introduce the PHI test with a threshold value of 35 after DRE and PSA testing in those men with normal DRE and a PSA level of 4-10 ng/ml as a rule-out test to avoid unnecessary biopsies if patients and clinicians were prepared to accept a risk of missing a high-grade cancer of less than 1%. Men with a PHI over 35 would continue to undergo TRUS biopsy and men below this level would undergo an annual PSA test until the age of 78 when mean survival from PCa exceeds life expectancy (Hamdy et al, 2016; Census and Statistics Department, Hong Kong SAR, September 2017).

The aim of economic evaluation is to consider the costs and consequences of a course of action in order to determine whether this use of resources is better than the next available alternative (Drummond et al, 2015). Although there would be additional costs due to the introduction of the test, short term savings would be made on biopsies avoided and the treatment of adverse events arising following biopsy. Health outcomes would be positively impacted for those

patients avoiding negative biopsies and its attendant complications. However, there would be an increased risk of missed cases of cancer in the group who avoid biopsy. This study presents a cost-consequence analysis of the introduction of the PHI test at PHI thresholds of 25, 35 and 55.

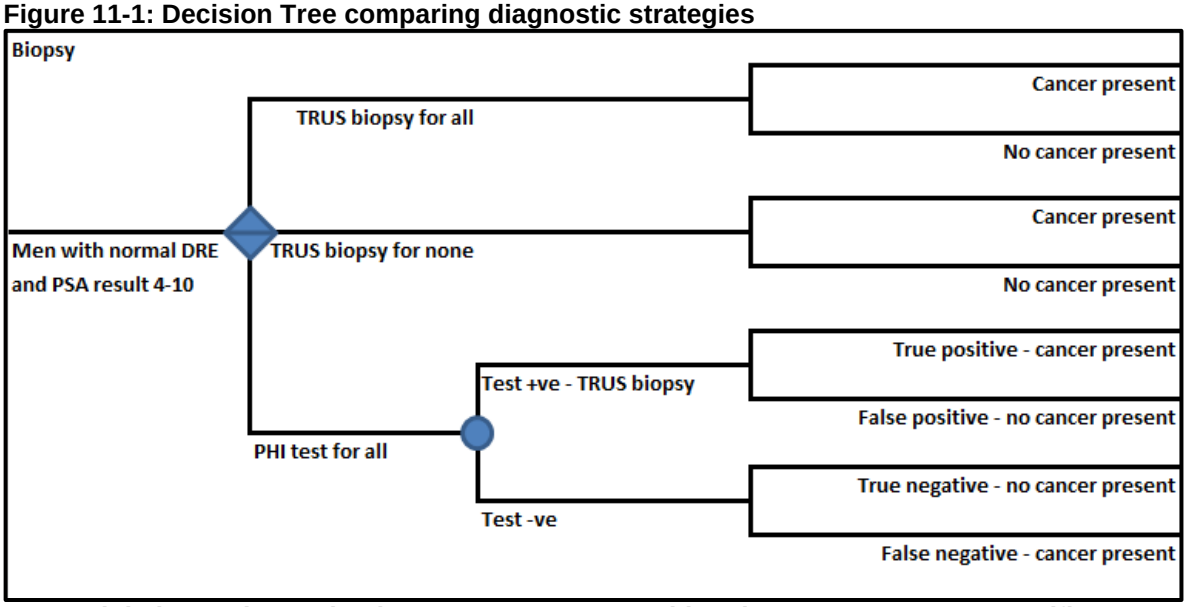
The current strategy (biopsy all) was mapped and two alternative diagnostic strategies (biopsy none and test then biopsy) for men with normal DRE and PSA levels 4-10 ng/ml using information supplied by Hong Kong clinicians (the decision-makers in this case study). For the test-then-biopsy strategy the costs and consequences were estimated for three different cut-off levels for the test. These pathways were set out as a decision tree with a time horizon covering the diagnostic process up to biopsy (see Figure 11-1). A Hong Kong public health service perspective was adopted. Health outcomes differ between the three strategies due to the direct impact of the biopsy and adverse effects following a proportion of biopsies and the proportion of cancer cases that are missed. Health outcomes relating to the biopsy procedure itself or complications of the procedure are not considered in this study but they would be positive under each of the testing strategies as they reduce as the number of biopsies reduces. The proportion of missed cancers which would be likely to result from each testing strategy was calculated and presented as well as the costs of the alternative strategies.

Figure 11-1 shows the current strategy under which all patients undergo TRUS biopsy, a strategy where no patients undergo biopsy and the proposed testing strategy. In the proposed pathway (the lowest branch of the decision tree) all patients are tested using PHI. Those with a positive result undergo TRUS biopsy. Those with PHI scores under the threshold (negative result) do not undergo biopsy. In both positive and negative test result arms there is the possibility of the test result being correct or incorrect resulting in missed cases of cancer. For each strategy cost savings were calculated on the basis of the number of biopsies avoided.

Clinical data to populate the decision trees and costing data for all costs apart from the PHI test were collected as part of the study reported by Chiu et al (2016). Sensitivity and specificity for three different thresholds of PHI score are shown in Figure 11-2. Any accident and emergency attendances and length of

hospital stay (where appropriate) following biopsy were recorded. This resource usage was valued using costs from the Annual Report for 2016-7 of the Hospital Authority Hong Kong (2017). The PHI test cost was based on the cost of the test at a Hong Kong clinic to a private user. Costs are not discounted given the short time horizon of the economic evaluation.

Figure 11-1: Decision Tree comparing diagnostic strategies



DRE –Digital Rectal Examination, PHI – Prostate Health Index, PSA - Prostate Specific Antigen, TRUS – Transrectal Ultrasound-guided biopsy. Diamond represents a decision node. Circle represents a probability node.

Figure 11-2: Test performance of PHI test at cut-off scores of 25, 35 and 55

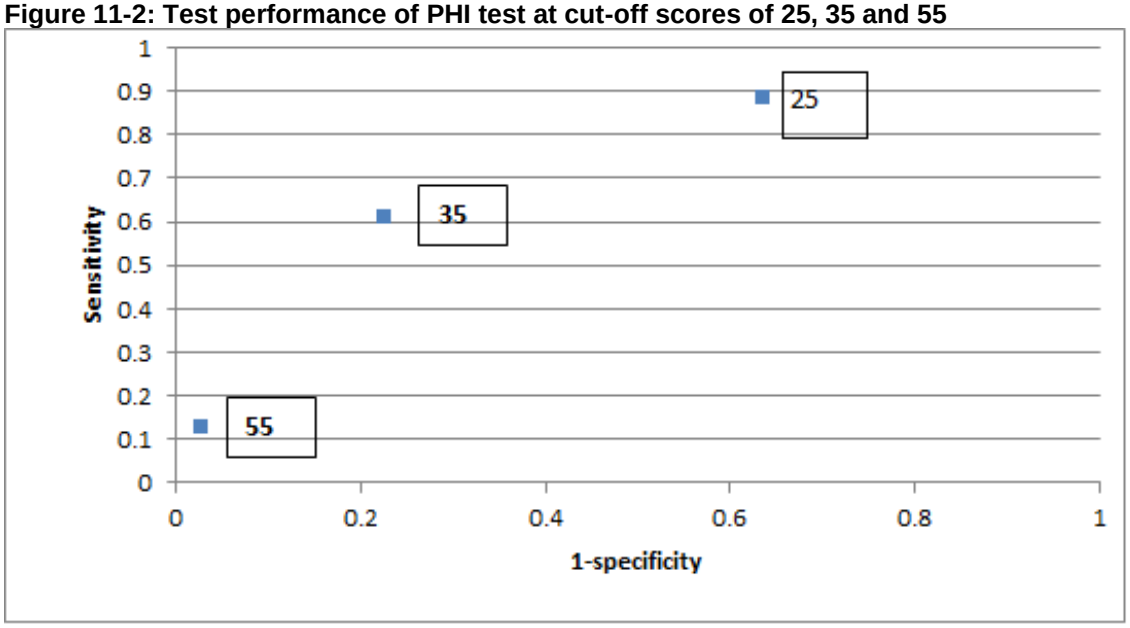


Table 11-1: Inputs to the model and data sources

Epidemiology and PHI test performance	
Prevalence	10.9%
Sensitivity of PHI test at cut-off 25 [1]	88.7%
Specificity of PHI test at cut-off 25 [1]	36.5%
Sensitivity of PHI test at cut-off 35 [1]	61.3%
Specificity of PHI test at cut-off 35 [1]	77.5%
Sensitivity of PHI test at cut-off 55 [1]	12.9%
Specificity of PHI test at cut-off 55 [1]	97.4%
Proportion of patients suffering an adverse event after biopsy [2]	0.07
Proportion of patients suffering an adverse event who require hospitalisation [2]	0.38
Costs	*HK$*
Cost of PHI test [3]	3,000
Cost of adverse event – Accident and Emergency Department Attendance [4]	1,300
Cost of adverse event - hospitalised [5]	23,116
Cost of TRUS biopsy [6]	10,900

Notes
[1] Data collected in Chiu at al study (2016)
[2] Unpublished data from study reported in Chiu et al (2016) - 39 patients attended Accident and Emergency Department from a cohort of 569 undergoing biopsy
[3] Based on the cost of the test to a private patient at a Hong Kong clinic
[4] Cost of Accident and Emergency Attendance from Annual Report 2016-7 Hospital Authority (2017)
[5] 4.67 days - mean length of stay for patients requiring hospitalisation (Data collected in Chiu et al study (2016) these data unpublished). Valued at HK$4,950 per day Accident and Emergency Attendance from Annual Report 2016-7 Hospital Authority (2017)
[6] Cost of TRUS biopsy based on hospital finance department analysis (unpublished)

HK$ - Hong Kong dollars, PHI – Prostate Health Index, TRUS – Transrectal Ultrasound-guided

One-way sensitivity analysis was undertaken whereby input parameters were varied in turn by 50% of base case value in both directions (whilst all other inputs to the model were held constant) to determine the impact of an over or under-estimation on the base-case results. This form of sensitivity analysis was undertaken as this allowed decision makers to assess the individual impact of each of the input parameters. Where sensitivity analysis showed that the result was sensitive to an individual parameter threshold analysis was undertaken.

This involved varying the range to determine the highest level at which the cost savings would be reduced to zero.

A total of 569 patients were included in the study by Chiu et al (2016). 62 of them were diagnosed with prostate cancer of which 16 were high grade cancers. Using a cut-off of 35, the introduction of the PHI test into the diagnostic pathway for men with normal DRE and PSA levels 4-10 ng/ml would save HK$5,500 per patient (see **Table** 11-1). At this cut-off the study data (Chiu et al, 2016) indicated that cancer in 4.2% of the population (including 0.53% with high grade cancers) may be missed. The majority of the saving results from approximately 75% of patients avoiding TRUS biopsy as their PHI score was under 35. A further cost saving of HK$511 results from a reduction in adverse events following biopsies. If the cut-off for PHI testing were increased to 55 over 95% of biopsies could be avoided resulting in an overall cost saving estimated to be in excess of HK$8,000. However, applying a cut-off of 55 to study data, prostate cancer in 9.5% of the population would have been missed including 2.1% with high grade cancer. If the cut-off is reduced to 25 around a third of biopsies could be avoided with just over 1% of all cancer cases missed including less than 0.2% of high-grade cancers. At this cut-off level cost savings are reduced to HK$914. At all cut-off levels the introduction of PHI results in cost savings although these are greater at higher cut-off levels as more biopsies are avoided. Testing costs for PHI are included in the analysis at HK$3,000 and it is proposed to test all patients in this population. Testing costs may reduce under the new testing strategy as a result of increased volumes, but this has not been reflected in the analysis

Table 11-2: Cost-consequence analysis of alternative diagnostic strategies

Strategy	Biopsy rate	Cost of biopsies (HK$)	Cost of PHI test (HK$)	Cost of adverse events (HK$)	Total cost (HK$)	Cost savings compared to biopsy all	Missed cancer cases - all	Missed cancer cases - high grade Gleason 7 or above
Biopsy all	100%	10,900	0	698	11,598	-	-	-
PHI test for all - cut off 25	66.26%	7,222	3,000	463	10,685	-914	1.23%	0.18%
PHI test for all - cut off 35	26.71%	2,912	3,000	187	6,098	-5,500	4.22%	0.53%
PHI test for all - cut off 55	3.69%	402	3,000	26	3,428	-8,170	9.49%	2.11%
Biopsy none	0%	0	0	0	0	-11,598	10.90%	2.81%

PHI – Prostate Health Index, HK$ - Hong Kong dollars

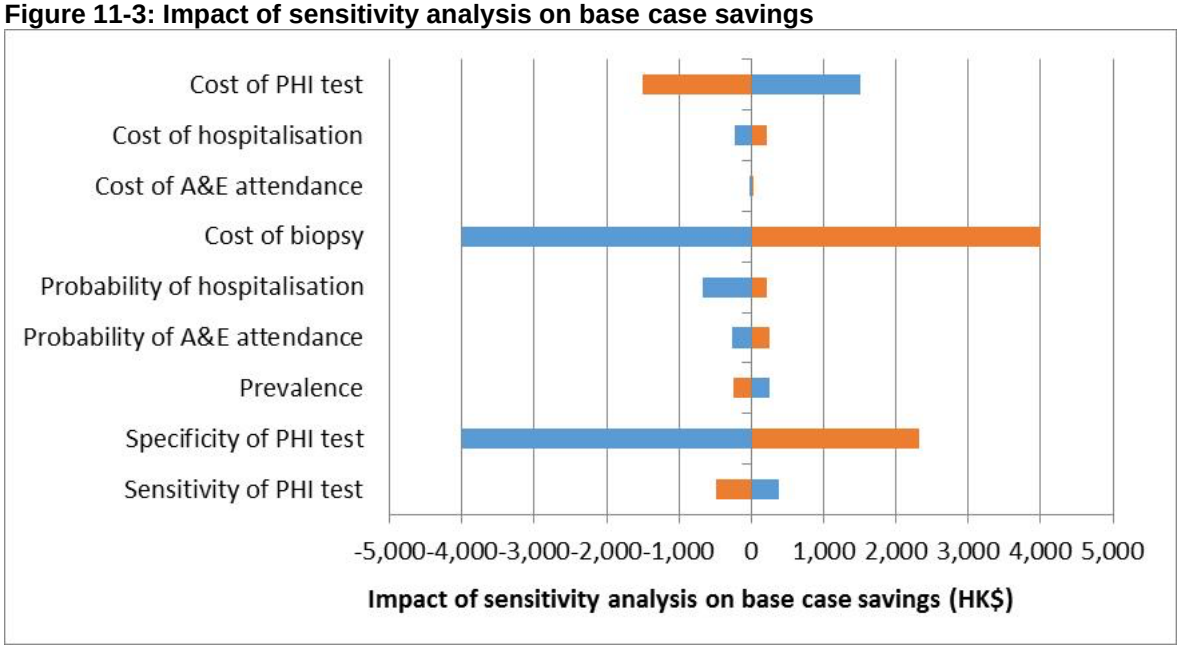
Sensitivity analysis showed that the overall cost saving was sensitive to the cost of biopsy, the cost of the PHI test and the specificity of the PHI test (Figure 11-3). No individual parameter, when varied within a range 50% above or below the base case would alter the conclusion that the introduction of the test is likely to be cost saving. Threshold analysis for these parameters determined values at which the proposed strategy would be cost neutral. These were 24% for the specificity of the PHI test (base case 77.5%), HK$3,400 for cost of biopsy (base case HK$10,900) and HK$8,500 for the cost of the test (base case HK$3,000). Table 11-3 summarises the base case results and the results of sensitivity and threshold analyses. Reducing the sensitivity of the PHI test whilst holding specificity constant results in more missed cases (all grades of cancer in 7.2% of the population missed at 34% sensitivity, compared to 4.2% in the base case of 61%) but increases cost savings as more biopsies are avoided. Lower specificity (with constant sensitivity) results in the same level of missed cases but savings are reduced as less biopsies are avoided.

Table 11-3: Results, sensitivity and threshold analysis

Base case (PHI cut-off 35)					
Cancer cases missed – all grades (high grade)			n=24/569 (3/569) or 4.2% (0.5%)		
Cost saving per patient					HK$5,500
Made up of:					
Additional costs of testing					(HK$3,000)
Direct cost savings from biopsies (circa 73% of patients at HK$10,900)					HK$7,988
Cost savings from reduction in adverse events					HK$511
Sensitivity analysis	Base case	Range (+/- 50%)		Results of sensitivity analysis (HK$)	
		Lower	Upper	Lower	Upper
Sensitivity of PHI test	61%	31%	100%	5,888	5,011
Specificity of PHI test	78%	39%	100%	1,499	7,824
Prevalence of all grades of cancer	11%	5%	16%	5,748	5,257
Proportion of patients experiencing adverse events	7%	3%	10%	5,242	5,757
Proportion of patients with adverse event requiring hospital	38%	19%	58%	4,831	5,724
Proposed test costs (HK$)	3,000	1,500	4,500	7,000	4,000
Cost of biopsy (HK$)	10,900	5,450	16,350	1,506	9,494
Costs of adverse events without hospitalisation (HK$)	1,300	650	1,950	5,467	5,533
Costs of adverse events with hospitalisation (HK$)	23,116	11,558	34,674	5,277	5,723
Threshold analysis	Base case	Value for proposed strategy to be cost neutral			
Specificity of PHI test (at 61% sensitivity)	78%				24%
Proposed test costs (HK$)	3,000				8,500
Cost of biopsy (HK$)	10,900				3,400

HK$ - Hong Kong Dollars, PHI – Prostate Health Index

Figure 11-3: Impact of sensitivity analysis on base case savings



A&E- accident and emergency department, HK$ - Hong Kong dollars, PHI – Prostate Health Index

11.4 Discussion

The study demonstrates the use of simple economic evaluation in a preliminary assessment of a diagnostic technology using local data. The study was relatively quick and resource-light as a result of the simplicity of the model and the availability of locally relevant data from a previous study. Evidence was taken from a single clinical study (Chiu et al, 2016) and micro-costings from a single hospital. One-way sensitivity analysis was appropriate as it allowed decision makers to assess the importance of individual parameters. A simplifying assumption was made that all men in the population currently undergo TRUS biopsy.

A significant limitation of this study is that data have been taken from a retrospective analysis of a cohort taken from a single clinical trial. A further limitation is that the model does not take account if a proportion of cases where patients and clinicians decide that biopsy is not their preferred option (in the current diagnostic strategy). In order to change the conclusion of the base case analysis (at a PHI cut-off of 35) just under 50% of men would need to refuse biopsy. The proportion of men not undergoing biopsy is believed to be substantially lower than this.

This study found that the adoption of the PHI test for patients with negative DRE and PSA score 4-10 ng/ml in Hong Kong has the potential to deliver significant cost savings although there are implications with a proportion of all grade cancers missed. The cost savings arise because the PHI test stratifies men into those requiring TRUS biopsy and those who can avoid biopsy and enter the monitoring programme. As biopsy and the adverse events associated with it are expensive to deal with, a strategy which avoids a relatively small proportion of biopsies has the potential to deliver savings which exceed the costs of testing all the patients in this population but this must be balanced with the risks of missed cases and the longer term cost and clinical outcomes.

The results are consistent with three previous economic evaluations of PHI. The first study by Nichol et al (2011) was a budget impact analysis of PHI plus total PSA and percent free PSA compared to PSA alone. This study evaluated the impact on 1-year total costs of PHI plus PSA to PSA alone in a screening programme from a US societal perspective in men 50-75 years old. Using thresholds for PHI testing of 2ng/ml and 4ng/ml they estimated cost savings of US$356,647 and US$94,219 respectively in a notional insurance company cohort of 100,000 men. 90% of the overall savings came from avoiding unnecessary biopsies. A further study by Nichol et al (2012) extended their previous analysis to a cost-utility analysis with a 25-year time horizon. This extended analysis found that PHI plus PSA dominated PSA alone strategy for both 2ng/ml and 4ng/ml thresholds delivering cost savings of US$1,199 and US$443 respectively together with utility gains of 0.08 and 0.03. Both Nicholl et al studies used data relevant to the US population and are not directly applicable to a Chinese population. The final economic evaluation study identified was Heijnsdijk et al (2016) who assessed the cost-effectiveness of using a PHI cut-off of 25 as an add-on to PSA with a cut-off of 3ng/ml in a European screening population aged 50-75. This study found a reduction in negative biopsies of 23%, a reduction of 17% in costs of diagnosis and 1% in total cost of prostate cancer. Although these results support the results of the study care must be taken as the clinical and cost-effectiveness of PHI in the Hong Kong setting is dependent upon epidemiology specific to the Chinese population as well as local treatment pathways, the organisation of local services and the capacity and organisation of molecular pathology services.

This case-study was a preliminary cost-consequence analysis that indicates that the PHI test is potentially cost-effective in that it is not "dominated" by current practice (i.e. costs more with worse outcomes). This represents necessary but not sufficient condition for cost-effectiveness. Further research is needed to compare the potential long-term cost and clinical consequences of missed cancer diagnoses against the short-term benefits.

Nicholl at al (2011) suggests that the negative consequences of missed cases are limited as they are likely to be found in subsequent screenings (Crawford and Abrahamsson, 2008). Effectively false negatives represent delayed rather than missed diagnoses. Moreover, cancers missed tend to have relatively-low Gleason scores and most cancers found 2-4 years after an initial screen are still curable (Wolters et al, 2010; Hugosson et al, 2003; van der Cruijsen-Koeter et al, 2003; Hoedemaeker et al, 2001; Postma et al, 2004; Hugosson et al, 2004; Schroder et al, 2008). However, the extent to which these conclusions hold in a Chinese population requires further study.

The immediate implication of this study for policy-makers is that in the Hong Kong context, PHI could be a cost-saving addition to the diagnostic set-up for prostate cancer in men with PSA levels of 4-10ng/ml and negative DRE. Although health outcomes have not been fully quantified, the analysis suggests that, at the proposed cut off of 35, sensitivity could be retained such that all grade cancer would missed in only 4.2% of the population and high-grade cancer in only 0.53% whilst a high proportion of biopsies would be avoided. The use of the PHI test in Hong Kong appears to warrant further investigation.

The decisions to be informed were whether further research would appear to be justified and if so, what form should that research take. Moreover, the study aimed to inform reimbursement strategy for the development team. As the study suggested that the introduction of PHI into the diagnostic pathway in this population had the potential to deliver significant savings then it would appear that further research would be justified. The current study suggests that results are sensitive to test performance and the costs of the biopsy and the test itself. The cost of the biopsy is known from a micro-costing exercise so is unlikely to vary to the extent that it has been varied in the present study. The cost of the test can be established once volumes are known and through further dialogue

with the suppliers. The key parameters for future research are, therefore, sensitivity and specificity of the test in the relevant population. The development team should also continue dialogue with the Hong Kong Health Authority, patients and clinicians to determine their views of an acceptable level of missed cases to ensure that a prospective study is designed at a cut-off which is acceptable to key stakeholders. This dialogue with the ultimate decision-maker (the Hong Kong Health Authority) also ensures that the study is designed to meet their evidence requirements. A useful design would be a randomised controlled trial where one arm of the study has the PHI test and one arm does not. Both arms would continue on to biopsy but clinicians would have an opportunity to state whether or not they would have not biopsied. A study type which would be useful would be a feasibility study which examined practical issues such as likely volumes, budget and staffing implications, how samples would be handled and time taken for turn-around. A resource analysis would also be useful as this would build-in the impact of the change in diagnostic pathway on the existing pathway as the introduction of the test would reduce numbers requiring biopsy, thus demand on TRUS facilities. Unless these facilities (staff, equipment, space) were able to be redeployed in the short term there may be some duplication of cost.

12 Discussion

12.1 Introduction

This chapter summarises the main findings, contributions and limitations of the thesis (Section 12.2, 12.3 and 12.4). It also suggests what the implications of the research are and sets out some recommendations for future research (sections 12.5 and 12.6). It addresses research questions 5 and 6.

5. What are the wider implications of the results of the study?

6. What recommendations for policy and future research arise from the study?

12.2 Main findings of the thesis

12.2.1 Features of DF-HTA

A list of ten features characterising DF-HTA was proposed. Four of the features (target audience, decisions to inform, available evidence and timing) had been included in previous frameworks distinguishing early and mainstream HTA (Ijzerman and Steuten, 2011) or classical HTA (Pietzsch and Paté-Cornell, 2008). A further six features were added (underlying user objective, decision space, business model, resources available for analysis, stance of analysis and burden of proof). The last two features are critical to understanding the particular nature of DF-HTA. The analyst takes a positive stance of analysis, effectively putting him or herself into the shoes of the developer to maximise the potential of the technology. This is particularly evident as much development is 'technology-driven' i.e. the invention precedes the identification of the precise clinical need. The first task of the analyst is to 'position' the technology in the most favourable clinical indication and place in the care pathway. This may involve considering multiple potential positions in what has been termed an 'exploratory analysis'. In DF-HTA the burden of proof is not established by any external body in the sense that there is no methods guidance and no pre-set criteria to meet. The analyst and developer need to consider the decisions to be made and the evidence and resources available in planning the analysis. Any limitations in the

evidence available or the methods adopted are discussed with the developers and limitations addressed, if appropriate, in later development stages.

12.2.2 Process of DF-HTA

A generic process of DF-HTA was proposed identifying clinical value assessment and economic value assessment as core activities to be undertaken iteratively as a development proceeds. Research and development and other commercial activities were recognised as closely linked but distinct activities. Information flows between each of the activities were identified. Both research and development and other commercial activities inform and are informed by clinical and economic value assessments.

12.2.3 Methods of DF-HTA

Clarity in the process of DF-HTA is useful to 'ring fence' methods available to the analyst separating assessment methods from methods used in research and development and other commercial activities. Methods identified as useful for DF-HTA were care pathway analysis, qualitative methods of stakeholder consultation, literature reviews, multi-criteria decision analysis, discrete choice experiments, expert opinion and expert elicitation, cost-effectiveness analysis and value of information analysis. For DF-HTA where resources are limited, less formal versions of the methods may be appropriate such as targeted literature reviews rather than systematic literature reviews. Cost effectiveness analysis may take the form of cost minimisation analysis or cost consequence analysis rather than cost utility analysis. Simpler modelling techniques, intermediate outcomes and short time horizons may be appropriate in DF-HTA. Estimands such as headroom and the thresholds for costs and clinical effectiveness (including test performance) required for a technology to be cost-effective in the modelled jurisdiction may also be informative in DF-HTA. There is some debate in the academic literature at present about the role of probabilistic sensitivity analysis and value of information analysis in early health economic modelling. Although there may be a role for these methods later in the development process, simpler approaches appear more appropriate for SME and academic developers working outside translational research bodies at early stages of technology development.

12.2.4 Illustration of framework using case studies

The list of features did appear to be able to distinguish DF-HTA when the case studies were compared against the list of features. There is an interesting contrast between the three case studies where the features of DF-HTA set out in the list are evident and the two case studies where this is not the case. For example, the case study concerning the melanoma test is one which demonstrated all the features in the framework. The cost-consequence analysis of the Prostate Health Index, a commercially available test shared three features with DF-HTA: burden of proof (as no specific methodological guidance for health economic analysis was followed); limited evidence; and constrained resources. However, in other features the study differed from DF-HTA. The audience included a public payer, the timing was late as the technology was commercially available; and the decision space was limited as the indication and position in the clinical pathway was known. Although the methods used were similar to those which may be used in DF-HTA (deterministic modelling, use of cost consequence analysis) this was primarily driven by the lack of resources available.

12.3 Contributions of the thesis

The first contribution of the thesis is the development of a framework of DF-HTA including features, process and methods. The list of features builds on the work of Pietzsch and Pate Cornell (2008) and Ijzerman and Steuten (2011) who set out frameworks of features distinguishing early HTA from classical or mainstream HTA. This thesis makes the argument that the needs of the audience (the developer) drive many of the distinct characteristics of the analysis, rather than the timing of the analysis. The proposed generic process of DF-HTA builds on the work of Cosh et al (2007) and Markiewicz et al (2014, 2017a). Cosh et al (2007) set out a sequential process for investment decisions for new technologies. Markiewicz et al (2014) classified methods under a number of aims which were later termed ‘areas of early assessment’ (Markiewicz et al, 2017a). This thesis offers a reconfiguration of Markiewicz et al’s ‘areas of early assessment’ as an iterative process of clinical and economic value assessment. The summary of suitable methods for DF-HTA built on the work of Markiewicz et al (2014) who presented a comprehensive synthesis of methods used in empirical

studies from a systematic review of applied studies. The specific contribution of this thesis is to simplify this synthesis by including similar methods under one name and excluding methods of research and development and other commercial activities. Building on the work of Vallejo-Torres et al (2008) and Chapman (2013), adaptations to methods were suggested for situations where resources are constrained.

The second contribution is to provide five empirical examples, three in DF-HTA and two in early economic evaluation. All concern diagnostic technologies. A number of recent studies have identified the need for development of early economic evaluation methods in addressing the potential value of stratified medicine approaches (Faulkner et al, 2012; Gaultney, 2014; Ijzerman et al, 2017). The case studies do not directly apply the features, process and methods identified in the methodological chapters as the case studies were completed at various different time points during the iterative development of the methodological work. They do, however, provide an opportunity to illustrate the framework of features, process and methods proposed.

12.4 Limitations

12.4.1 Literature search

The main limitation of any review of HTA to inform developers is that the majority of work in this area is not published due to commercial sensitivity (Tu et al, 2014). The published studies may represent those where methods applied were more complex and/or where more resources were available for analysis. The search strategy may also not have identified all relevant examples of applied or methodological studies of development-focused HTA. As the aim of the search was to identify features and useful methods of DF-HTA any omission would only be a concern if it discussed or applied a new methodological approach not discussed anywhere in the identified studies. Screening of study titles and data extraction was done by a single researcher who also decided when saturation point had been reached on methodological points. It is probable that a subsequent researcher would select a different set of studies from the ones reviewed in this thesis. Again, this would only be an issue of

concern if a significant area of interest was not represented or was misrepresented.

12.4.2 Subjectivity in paper selection and framework development

As the audience in many studies was not made explicit it was difficult to separate those studies where the intended audience was the developer from other early HTA studies. The judgement was made by one researcher with discussion with supervisors in cases of uncertainty. The inclusion criteria were that the audience was explicitly developers or that some reference was made to decisions which were development-focused (such as technology design). These inclusion criteria may have resulted in a smaller pool of applied DF-HTA studies as some DF-HTA studies may have been excluded.

Features of DF-HTA were extracted from a mix of studies with explicit and implicit frameworks. Where the framework was implicit, both the selection of the studies to include and the extraction of the features were subjective. This would only have an impact where the selection of a different paper would have called into question the inclusion of one of the features or added a further feature. The validation of the framework in future studies by other research groups would mitigate this risk.

12.4.3 Lack of formal methodology and validation

Development of the framework of features, process and methods of DF-HTA involved iterative processes of literature searching, discussion and drafting between myself and the supervisors of the thesis. No formal methodology was used although this process shared some features with framework analysis (Spencer and Ritchie, 2002; Oliver et al, 2008). A recent study suggests that this method of framework development is not unusual, but a more formal process of framework development would have made the steps more transparent for the reader and may have improved internal validity (McMeekin et al, 2020). No external validation has been carried out to date. The validation of the framework by other research groups would be useful to establish external validity.

12.5 Implications

12.5.1 Implications for developers

Evidence suggests that many developers of health technologies do not use formal methods of decision-support and do not have in-house capacity and knowledge to perform HTA (Craven et al, 2012; Markiewicz et al, 2017a). The case studies in this thesis suggest that there are benefits for developers in consulting an HTA practitioner at an early stage in the development process. HTA is able to inform a go/no go decision at any point in development. In a 'needs driven' scenario such as the ovarian cancer case study, HTA methods can be used to estimate whether there is potential value in pursuing the project. This evaluation of the 'room for improvement' can take place without any significant resource commitment on the potential developers' part. Where development is 'technology-driven' it is beneficial to start to 'position' the technology and articulate value propositions for the technology in the most promising positions in the care pathway. A key aspect of DF-HTA at the earliest stages in development is to encourage engagement with potential users and other stakeholders as this may impact on technology design and evidence generation strategy. For example, in the melanoma case study, the potential cost savings available could be undermined if GPs were risk averse and referred all patients with a mole that appeared suspicious using their clinical judgement as well as those identified as needing referral according to the new test. Evidence would be required to determine the level of GPs' risk aversion.

Two of the case studies contribute generic tools which developers could use without the involvement of an HTA practitioner. The Test and Treat Superiority Plot and model (Chapter 7) would be useful for developers of treatment response tests as a first hurdle. The model presented in Chapter 10 is a flexible tool for developers of potential triage tests with application whether or not the reference testing capacity is constrained. Often, developers and investors find aspects of test performance difficult to communicate. These simple models allow parties to explore the impact of different levels of test performance in a simple format.

Discussions with an HTA practitioner may also encourage developers to keep in mind the breadth of indications, positions in the care pathway, jurisdictions, business models and jurisdictions which may be available to them.

12.5.2 Implications for practitioners of HTA

As DF-HTA is a relatively new field, practitioners who have previously concentrated on HTA undertaken to support reimbursement or adoption decisions may find the framework set out in this thesis a useful introduction to the field. The practitioner could consider the table of features to clarify the nature of the study which, if reported, would improve clarity for readers. The generic process (section 5.3.1.2) provides clarity about the role of DF-HTA in relation to research and development and other commercial activities. Practitioners may find it useful to undertake clinical value assessment as a separate exercise to economic value assessment early in the development process. Undertaking care pathway analysis, modelling of clinical outcomes and/or engaging with stakeholders may be sufficient to inform developers' decisions at the earliest stage without requiring formal economic analysis.

HTA practitioners familiar with reimbursement-focused HTA may look to undertake a full cost-utility analysis if the evidence is available to complete it. It may, however, be a better use of the resources available to use a simpler model and use the resource available to conduct, for example, stakeholder consultation to explore likely behavioural issues. The purpose of the HTA through the development process is to inform ongoing decision-making rather than provide a one-off recommendation. Simple quantitative models are useful as a springboard for discussion.

Where a simple quantitative exercise is undertaken, for example, a headroom analysis, it is important to keep in mind that this has narrowed the focus to a particular jurisdiction, indication and business model. It is often necessary to narrow down, for practical reasons, but a technology should not be rejected before exploring the full breadth of possibilities in terms of indications, positions in pathway, jurisdictions and business models.

12.5.3 Implications for policy-makers

DF-HTA can be used to prioritise research either on a portfolio basis or by assessing projects on a stand-alone basis. Policy-makers should encourage the robust articulation of a value proposition (particularly for translational research) within funding applications. Lehoux et al (2017) highlight the disconnect between the innovation funding stream and health research funding streams. In particular, they note that innovation funding may not give sufficient focus as to whether a potential health innovation is relevant, usable and sustainable. DF-HTA may offer an opportunity to align the two agendas.

The case studies have exemplified the difficulties in translating molecular pathology and other precision medicine technologies. For example, the melanoma test illustrated the difficulty in generating context-specific evidence at a reasonable cost. Translational research bodies which facilitate links between commercial entities, academic and clinical researchers, pathologists and clinicians, regulators and reimbursement agencies may improve the translation rate. Policy-makers should continue to fund translational research bodies.

Full disease models, such as that developed by the Innovation and Value Initiative could greatly improve the efficiency of DF-HTA (The Value Initiative, 2020). This may have allowed the development of a lifetime horizon in the rheumatoid arthritis case study. Such full-disease models in priority areas could be used to evaluate any proposed technology (in technology-driven development) as well as determining areas of greatest need to inform specific calls for innovation (needs-driven development).

12.6 Recommendations for future research

Two of the case studies highlighted the use of similar methods to DF-HTA to inform decision-makers about the costs and consequences of expanding an existing molecular pathology test or introducing a new triage test into the clinical pathway. The decision-making process for the adoption of molecular pathology technologies is not clear and transparent and research into the evidence decision-makers rely on would be useful.

Publication of further empirical examples of HTA to inform developers would be useful, particularly if reporting of the features of the studies was explicit. Research exploring the features of in-house and unpublished DF-HTA would be useful as well as research exploring the usefulness of DF-HTA to developers.

This thesis has suggested a framework for DF-HTA comprising features, a generic process and analytic methods. Research into the usefulness and ease of use of the framework would be of benefit. A Knowledge Transfer Partnership has been funded by Innovate UK to transfer DF-HTA to an industry partner of the Glasgow Molecular Pathology Node and this will provide an opportunity to validate the approach.

The extent to which full disease models have been or could be used to assess the value proposition for innovative technologies would be a useful area of research. This may also allow a more needs-driven approach to innovation funding as called for by Lehoux et al (2017) and Greenhalgh et al (2018). Specifically, the models made available by the Innovation and Value Initiative (the valueinitiative, 2020) in non-small cell lung cancer and rheumatoid arthritis could be applied to relevant technologies in development to assess the feasibility of using this resource.

List of References

ABEL, L., SHINKINS, B., SMITH, A., SUTTON, A.J., SAGOO, G.S., UCHEGBU, I., ALLEN, A.J., GRAZIADIO, S., MOLONEY, E., YANG, Y. AND HALL, P., 2019. Early economic evaluation of diagnostic technologies: experiences of the NIHR diagnostic evidence co-operatives. Medical Decision Making, 39(7), pp.857-866.

ABPI. Undated. Top Pharmaceutical Markets by Value. [online] Available at: https://www.abpi.org.uk/facts-and-figures/global-pharmaceutical-market/top-10-pharmaceutical-markets-by-value-usd/. [Accessed 31 July 2020].

ACADEMY OF MEDICAL SCIENCES. 2016. Health Economics and stratified medicine. [online] Available at: https://acmedsci.ac.uk/file-download/61141574. [Accessed 31 July 2020].

ADAMS, E.R., AUGUSTIN, Y., BYRNE, R.L. et al. Rapid development of COVID-19 rapid diagnostics for low resource settings: accelerating delivery through transparency, responsiveness, and open collaboration (Preprint). Available at: https://doi.org/10.1101/2020.04.29.20082099 [Accessed 31 July 2020].

ANNEMANS, L., GENESTÉ, B. & JOLAIN, B. 2000. Early modelling for assessing health and economic outcomes of drug therapy. Value in Health, 3, 427-434.

ANSARI, M., 2013. The regulation of companion diagnostics: a global perspective. Therapeutic Innovation & Regulatory Science, 47(4), pp.405-415.

BA'PHAM, BAGHERI, E., RIOS, P., POURMASOUMI, A., ROBSON, R.C., HWEE, J., ISARANUWATCHAI, W., DARVESH, N., PAGE, M.J. AND TRICCO, A.C. 2018. Improving the conduct of systematic reviews: a process mining perspective. Journal of Clinical Epidemiology, 103, pp.101-111.

BARTELMES, M., NEUMANN, U., LÜHMANN, D., SCHÖNERMARK, M. P. & HAGEN, A. 2009. Methods for assessment of innovative medical technologies during early stages of development. GMS Health Technology Assessment, 5.

BAEYENS, J.P., SERRIEN, B., GOOSSENS, M. AND CLIJSEN, R., 2019. Questioning the “SPIN and SNOUT” rule in clinical testing. Archives of Physiotherapy, 9(1), pp.1-6.

BOELL, S. K. & CECEZ-KECMANOVIC, D. 2010. Literature reviews and the hermeneutic circle. Australian Academic & Research Libraries, 41, 129-144.

BOJKE, L., SOARES, M.F.O., FOX, A., JANKOVIC, D., CLAXTON, K.P., MORTON, A., SHARPLES, L., JACKSON, C.H., TAYLOR, A.C. AND COLSON, A., 2019. Developing a reference protocol for expert elicitation in healthcare decision making. Health Technology Assessment Reports.

BORSCI, S., BUCKLE, P., UCHEGBU, I., NI, M., WALNE, S. & HANNA, G. B. 2017. Integrating human factors and health economics to inform the design of medical device: a conceptual framework. EMBEC & NBC 2017. Springer.

BOSSUYT PM, IRWIG L, CRAIG J, GLASZIOU P. Comparative accuracy: assessing new tests against existing diagnostic pathways. BMJ. 2006 May 4;332(7549):1089-92.

BOYD, K. A. 2012. Employing early decision analytic modelling to inform economic evaluation in health care: theory & practice. PhD Thesis. University of Glasgow.

BRANDES, A., SINNER, M. F., KÄÄB, S. & ROGOWSKI, W. H. 2015. Early decision-analytic modeling–a case study on vascular closure devices. BMC Health Services Research, 15, 486.

BRETELER, M. 2012. Scenario Analysis and Real options modeling of home brain monitoring in epilepsy patients. Master’s Thesis, University of Twente.

BRIGGS, A., SCULPHER, M. AND CLAXTON, K., 2006. Decision modelling for health economic evaluation. OUP, Oxford.

BUISMAN, L. R., RUTTEN-VAN MÖLKEN, M. P., POSTMUS, D., LUIME, J. J., UYL-DE GROOT, C. A. & REDEKOP, W. K. 2016. The early bird catches the worm: early

cost-effectiveness analysis of new medical tests. International Journal of Technology Assessment in Health Care, 32, 46-53.

CADTH. 2017. Guidelines for the Economic Evaluation of Health Technologies: Canada – 4th edition. Available from https://www.cadth.ca/dv/guidelines-economic-evaluation-health-technologies-canada-4th-edition [Accessed 31 July 2020].

CAMPBELL, G. AND YUE, L.Q., 2016. Statistical innovations in the medical device world sparked by the FDA. Journal of Biopharmaceutical Statistics, 26(1), pp.3-16.

CAPDEVILLE, R., BUCHDUNGER, E., ZIMMERMANN, J. AND MATTER, A., 2002. Glivec (STI571, imatinib), a rationally developed, targeted anticancer drug. Nature Reviews Drug Discovery, 1(7), p.493.

CARAYON, P., WETTERNECK, T.B., RIVERA-RODRIGUEZ, A.J., HUNDT, A.S., HOONAKKER, P., HOLDEN, R. AND GURSES, A.P., 2014. Human factors systems approach to healthcare quality and patient safety. Applied Ergonomics, 45(1), pp.14-25.

CARLSSON, S., AUS, G., WESSMAN, C. AND HUGOSSON, J., 2007. Anxiety associated with prostate cancer screening with special reference to men with a positive screening test (elevated PSA)–Results from a prospective, population-based, randomised study. European Journal of Cancer, 43(14), pp.2109-2116.

CASSIDY, J., CLARKE, S., DÍAZ-RUBIO, E., SCHEITHAUER, W., FIGER, A., WONG, R., KOSKI, S., LICHINITSER, M., YANG, T.-S. & RIVERA, F. 2008. Randomized phase III study of capecitabine plus oxaliplatin compared with fluorouracil/folinic acid plus oxaliplatin as first-line therapy for metastatic colorectal cancer. Journal of Clinical Oncology, 26, 2006-2012.

CCG. CANCER ASSESSMENT 2017/18 Independent Panel Commentary Publications Gateway Reference: 08372. Available from: https://www.england.nhs.uk/wp-content/uploads/2018/08/ccg-cancer-assessment-2017-18.pdf [Accessed 31 July 2020].

CENTER FOR MEDICARE AND MEDICAID SERVICES. 2020. CLINICAL LABORATORY IMPROVEMENT AMENDMENTS. Available from: https://www.cms.gov/Regulations-and-Guidance/Legislation/CLIA [Accessed 31 July 2020].

CHAPMAN, A. M. 2013. The use of early economic evaluation to inform medical device decisions: an evaluation of the Headroom method. PhD Thesis, University of Birmingham.

CHEN, R., REN, S., YIU, M. K., FAI, N. C., CHENG, W. S., IAN, L. H., NAITO, S., MATSUDA, T., KEHINDE, E. & KURAL, A. 2014. Prostate cancer in Asia: a collaborative report. Asian Journal of Urology, 1, 15-29.

CHESBROUGH, H. 2006. Open Business Models: How to Thrive in the New Innovation Landscape, Boston, MA: Harvard Business School Press.

CHIU, P. K., ROOBOL, M. J., TEOH, J. Y., LEE, W.-M., YIP, S.-Y., HOU, S.-M., BANGMA, C. H. & NG, C.-F. 2016. Prostate health index (PHI) and prostate-specific antigen (PSA) predictive models for prostate cancer in the Chinese population and the role of digital rectal examination-estimated prostate volume. International Urology and Nephrology, 48, 1631-1637.

CLARK, J., GLASZIOU, P., DEL MAR, C., BANNACH-BROWN, A., STEHLIK, P. AND SCOTT, A.M., 2020. A full systematic review was completed in 2 weeks using automation tools: a case study. Journal of Clinical Epidemiology, 121, pp.81-90.

CLAXTON K, MARTIN S, SOARES M, et al. Methods for the estimation of the National Institute for Health and Care Excellence cost-effectiveness threshold. Southampton (UK): NIHR Journals Library; 2015 Feb. (Health Technology Assessment, No. 19.14.) Appendix 1, Systematic review of the literature on the cost-effectiveness threshold. Available from: https://www.ncbi.nlm.nih.gov/books/NBK274312/

CLINICAL TRIALS REGISTER. 2017. Available at https://www.clinicaltrialsregister.eu/ctr-search/trial/2017-000370-10/GB#A [Accessed 31 July 2020].

COCHRANE AL. 1973. Effectiveness and efficiency: random reflections on health services. London: Nuffield Provincial Hospitals Trust.

COSH, E., GIRLING, A., LILFORD, R., MCATEER, H. & YOUNG, T. 2007. Investing in new medical technologies: a decision framework. Journal of Commercial Biotechnology, 13, 263-271.

COX, N. 2004. Evaluation of the UK 2-week referral rule for skin cancer. British Journal of Dermatology, 150, 291-298.

CRAVEN, M. P., ALLSOP, M. J., MORGAN, S. P. & MARTIN, J. L. 2012. Engaging with economic evaluation methods: insights from small and medium enterprises in the UK medical devices industry after training workshops. Health Research Policy and Systems, 10, 29.

CRAVEN, M. P. & MORGAN, S. P. 2011. Early Stage Economic Evaluation with a Small Medical Device Start-up Company using a Markov Model. Transactions of the ASME-W-Journal of Medical Devices, 5, 027516.

CRAVEN, M. P., MORGAN, S. P., CROWE, J. A. & LU, B. 2009. Deploying a spreadsheet tool for early economic value assessment of medical device innovations with healthcare decision makers. Journal of Management & Marketing in Healthcare, 2, 278-292.

CRAWFORD, E. D. & ABRAHAMSSON, P.-A. 2008. PSA-based screening for prostate cancer: how does it compare with other cancer screening tests? European Urology, 54, 262-273.

CULYER, A.J., 2010. The dictionary of health economics. Edward Elgar Publishing.

CURTIS, L. A. & BURNS, A. 2018. Unit Costs of Health and Social Care 2017.

DAVEY, S. M., BRENNAN, M., MEENAN, B. J., MCADAM, R., GIRLING, A., CHAPMAN, A. & LILFORD, R. L. 2011. A framework to manage the early value

proposition of emerging healthcare technologies. Irish Journal of Management, 31, 59.

DE ANGELIS, R., SANT, M., COLEMAN, M.P., et al. Cancer survival in Europe 1999-2007 by country and age: results of EUROCARE-5 - a population-based study. Lancet Oncology. 2017;15(1):23-34. doi:10.1016/S1470-2045(13)70546-1.

DEGELING, K., KOFFIJBERG, H. AND IJZERMAN, M.J., 2017. A systematic review and checklist presenting the main challenges for health economic modeling in personalized medicine: towards implementing patient-level models. Expert Review of Pharmacoeconomics & Outcomes Research, 17(1), pp.17-25.

DE GRAAF, G., POSTMUS, D. & BUSKENS, E. 2015. Using multicriteria decision analysis to support research priority setting in biomedical translational research projects. BioMed Research International.

DE GRAAF, G., POSTMUS, D., WESTERINK, J. AND BUSKENS, E., 2018. The early economic evaluation of novel biomarkers to accelerate their translation into clinical applications. Cost Effectiveness and Resource Allocation, 16(1), p.23.

DENNY, L. Staging of ovarian cancer: time to subdivide more? Journal of Gynecologic Oncology. 2013;24(4):293-294. doi:10.3802/jgo.2013.24.4.293.

DI MASI, J.A., GRABOWSKI, H.G. AND HANSEN, R.W., 2016. Innovation in the pharmaceutical industry: new estimates of R&D costs. Journal of Health Economics, 47, pp.20-33.

DI RUFFANO, L.F., DINNES, J., TAYLOR-PHILLIPS, S., DAVENPORT, C., HYDE, C. AND DEEKS, J.J., 2017. Research waste in diagnostic trials: a methods review evaluating the reporting of test-treatment interventions. BMC Medical Research Methodology, 17(1), p.32.

DOBLE, B., 2016. Budget impact and cost-effectiveness: can we afford precision medicine in oncology? Scandinavian Journal of Clinical and Laboratory Investigation, 76(sup245), pp.S6-S11.

DONG, H. & BUXTON, M. 2006. Early assessment of the likely cost-effectiveness of a new technology: a Markov model with probabilistic sensitivity analysis of computer-assisted total knee replacement. International Journal of Technology Assessment in Health Care, 22, 191-202.

DOUMA KF, KARSENBERG K, HUMMEL MJ, BUENO-DE-MESQUITA, JM, VAN HARTEN WH: Methodology of constructive technology assessment in health care. International Journal of Technology Assessment in Health Care , 23(2):162-168, 2007.

DRANITSARIS, G. & LEUNG, P. 2004. Using decision modeling to determine pricing of new pharmaceuticals: the case of neurokinin-1 receptor antagonist antiemetics for cancer chemotherapy. International Journal of Technology Assessment in Health Care , 20, 289-295.

DRUMMOND, M.F., 2019. Modeling in Early Stages of Technology Development: Is an iterative approach needed?: Comment on “Problems and Promises of Health Technologies: The Role of Early Health Economic Modeling”. International Journal of Health Policy and Management.

DRUMMOND, M. F., SCULPHER, M. J., CLAXTON, K., STODDART, G. L. & TORRANCE, G. W. 2015. Methods for the economic evaluation of health care programmes, Oxford University Press.

DUCREUX, M., BENNOUNA, J., HEBBAR, M., YCHOU, M., LLEDO, G., CONROY, T., ADENIS, A., FAROUX, R., REBISCHUNG, C. & BERGOUGNOUX, L. 2011. Capecitabine plus oxaliplatin (XELOX) versus 5-fluorouracil/leucovorin plus oxaliplatin (FOLFOX-6) as first-line treatment for metastatic colorectal cancer. International Journal of Cancer, 128, 682-690.

EASTHAM, R. AND O’SHEA. Undated. Powerpoint presentation: The two-week wait suspected cancer referral process and teledermatology. [online] Available at: https://www.google.com/search?q=teledermatology+service+leeds&rlz=1C1GCEU_enGB821GB821&oq=tele&aqs=chrome.0.69i59j69i57j0l4.3035j1j8&sourceid=chrome&ie=UTF-8. [Accessed 31 July 2020].

EDDY, D. 2009. Health Technology Assessment and Evidence-Based Medicine: What Are We Talking About? Value in Health, 12, pp.S6-S7.

EISENBERG, R.S., 2015. Diagnostics need not apply. Boston University Journal of Science and Technology Law, 21, p.256.

EISENHAUER, E. A., THERASSE, P., BOGAERTS, J., SCHWARTZ, L. H., SARGENT, D., FORD, R., DANCEY, J., ARBUCK, S., GWYTHER, S. & MOONEY, M. 2009. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). European Journal of Cancer, 45, 228-247.

EU. 2017. REGULATIONS FOR MDR AND IVDR. Available from https://www.gov.uk/guidance/medical-devices-eu-regulations-for-mdr-and-ivdr [Accessed 31 July 2020].

EVANS, D., 2003. Hierarchy of evidence: a framework for ranking evidence evaluating healthcare interventions. Journal of Clinical Nursing, 12(1), pp.77-84.

FARGEN, K.M., FREI, D., FIORELLA, D., MCDOUGALL, C.G., MYERS, P.M., HIRSCH, J.A. AND MOCCO, J., 2013. The FDA approval process for medical devices: an inherently flawed system or a valuable pathway for innovation? Journal of Neurointerventional Surgery, 5(4), pp.269-275.

FARIA, R., SOARES, M.O., SPACKMAN, E., AHMED, H.U., BROWN, L.C., KAPLAN, R., EMBERTON, M. AND SCULPHER, M.J., 2018. Optimising the diagnosis of prostate cancer in the era of multiparametric magnetic resonance imaging: a cost-effectiveness analysis based on the Prostate MR Imaging Study (PROMIS). European Urology, 73(1), pp.23-30.

FAULKNER, E., ANNEMANS, L., GARRISON, L., HELFAND, M., HOLTORF, A.-P., HORNBERGER, J., HUGHES, D., LI, T., MALONE, D. & PAYNE, K. 2012. Challenges in the development and reimbursement of personalized medicine—payer and manufacturer perspectives and implications for health economics and outcomes research: a report of the ISPOR Personalized Medicine Special Interest Group. Value in Health, 15, 1162-1171.

FEDERICI, C. AND TORBICA, A. 2020. 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". International Journal of Health Policy and Management.

FRAPPIER, J., STANISH, W., BRITTBERG, M., STEINWACHS, M., CROWE, L., CASTELO, D. & RESTREPO, A. 2014. Economic evaluation of BST-CarGel as an adjunct to microfracture vs microfracture alone in knee cartilage surgery. Journal of Medical Economics, 17, 266-278.

FRUEH, F.W., 2013. Regulation, reimbursement, and the long road of implementation of personalized medicine—a perspective from the United States. Value in Health, 16(6), pp.S27-S31.

FUGEL, H.J., NUIJTEN, M., POSTMA, M. AND REDEKOP, K., 2016. Economic evaluation in stratified medicine: methodological issues and challenges. Frontiers in Pharmacology, 7, p.113.

GAULTNEY, J. G. 2014. Economic Evaluations of Targeted Therapy and Risk-Stratified Treatment Approaches in Multiple Myeloma. PhD Thesis, Erasmus University.

GIRLING, A., LILFORD, R., COLE, A. & YOUNG, T. 2015. Headroom approach to device development: current and future directions. International Journal of Technology Assessment in Health Care, 31, 331-338.

GIRLING, A., YOUNG, T., BROWN, C. & LILFORD, R. 2010. Early-stage valuation of medical devices: the role of developmental uncertainty. Value in Health, 13, 585-591.

GIRLING, A. J., FREEMAN, G., GORDON, J. P., POOLE-WILSON, P., SCOTT, D. A. & LILFORD, R. J. 2007. Modeling payback from research into the efficacy of left-ventricular assist devices as destination therapy. International Journal of Technology Assessment in Health Care, 23, 269-277.

GOODWIN P AND WRIGHT G. 2014. Decision analysis for management judgement. 5th edition. Wiley, Chichester, UK.

GOSLING, J. P. 2014. Methods for eliciting expert opinion to inform health technology assessment. Vignette Commissioned by the MRC Methodology Advisory Group. Medical Research Council (MRC) and National Institute for Health Research (NIHR).

GRANSTROM, C., SUNDQUIST, J., HEMMINKI, K. Population attributable fractions for ovarian cancer in Swedish women by morphological type. British Journal of Cancer. 2008;98(1):199-205. doi:10.1038/sj.bjc.6604135.

GRAZIADIO, S., WINTER, A., LENDREM, B.C., SUKLAN, J., JONES, W.S., URWIN, S.G., O'LEARY, R.A., DICKINSON, R., HALSTEAD, A., KUROWSKA, K. AND GREEN, K., 2020. How to Ease the Pain of Taking a Diagnostic Point of Care Test to the Market: A Framework for Evidence Development. Micromachines, 11(3), p.291.

GREENHALGH, T. AND SHAW, S., 2017. Understanding heart failure; explaining telehealth–a hermeneutic systematic review. BMC Cardiovascular Disorders, 17(1), p.156.

GREENHALGH, T., FAHY, N. AND SHAW, S., 2018. The Bright Elusive Butterfly of Value in Health Technology Development: Comment on" Providing Value to New Health Technology: The Early Contribution of Entrepreneurs, Investors, and Regulatory Agencies". International Journal of Health Policy and Management, 7(1), p.81.

GROOTHUIS-OUDSHOORN, C.G., FERMONT, J.M., VAN TIL, J.A. AND IJZERMAN, M.J., 2014. Public stated preferences and predicted uptake for genome-based colorectal cancer screening. BMC Medical Informatics and Decision Making, 14(1), p.18.

GRUTTERS, J. P., GOVERS, T., NIJBOER, J., TUMMERS, M., VAN DER WILT, G. J., & ROVERS, M. M. 2019. Problems and promises of health technologies: the role of early health economic modeling. International Journal of Health Policy and Management, 8(10), 575.

HAAKMA, W., STEUTEN, L. M., BOJKE, L. & IJZERMAN, M. J. 2014. Belief elicitation to populate health economic models of medical diagnostic devices in development. Applied Health Economics and Health Policy, 12, 327-334.

HAMDY, F. C., DONOVAN, J. L., LANE, J. A., MASON, M., METCALFE, C., HOLDING, P., DAVIS, M., PETERS, T. J., TURNER, E. L. & MARTIN, R. M. 2016. 10-year outcomes after monitoring, surgery, or radiotherapy for localized prostate cancer. New England Journal of Medicine, 375, 1415-1424.

HANSON, C.L., OSBERG, M., BROWN, J., DURHAM, G. AND CHIN, D.P., 2017. Conducting patient-pathway analysis to inform programming of tuberculosis services: methods. The Journal of Infectious Diseases, 216(suppl_7), pp.S679-S685.

HARTZ, S. & JOHN, J. 2008. Contribution of economic evaluation to decision making in early phases of product development: a methodological and empirical review. International Journal of Technology Assessment in Health Care, 24, 465-472.

HEIJNSDIJK, E. A., DENHAM, D. & DE KONING, H. J. 2016. The cost-effectiveness of prostate cancer detection with the use of prostate health index. Value in Health, 19, 153-157.

HJELMGREN, J., GHATNEKAR, O., REIMER, J., GRABOWSKI, M., LINDVALL, O., PERSSON, U. AND HAGELL, P., 2006. Estimating the value of novel interventions for Parkinson's disease: an early decision-making model with application to dopamine cell replacement. Parkinsonism & Related Disorders, 12(7), pp.443-452.

HOEDEMAEKER, R. F., VAN DER KWAST, T. H., BOER, R., DE KONING, H. J., ROOBOL, M., VIS, A. N. & SCHRØDER, F. H. 2001. Pathologic features of prostate cancer found at population-based screening with a four-year interval. Journal of the National Cancer Institute, 93, 1153-1158.

HOLFORD, N., MA, S.C. AND PLOEGER, B.A. 2010. Clinical trial simulation: a review. Clinical Pharmacology & Therapeutics, 88(2), pp.166-182.

HONG KONG HOSPITAL AUTHORITY. 2017. Annual Report 2016-7, available at https://www.ha.org.hk/visitor/ha_visitor_index.asp?Content_ID=240944&Lang=ENG [accessed 31 July 2020].

HUGHES, D.A., VILAR, F.J., WARD, C.C., ALFIREVIC, A., PARK, B.K. AND PIRMOHAMED, M., 2004. Cost-effectiveness analysis of HLA B* 5701 genotyping in preventing abacavir hypersensitivity. Pharmacogenetics and Genomics, 14(6), pp.335-342.

HUGOSSON, J., AUS, G., LILJA, H., LODDING, P., PIHL, C.G. AND PILEBLAD, E., 2003. Prostate specific antigen based biennial screening is sufficient to detect almost all prostate cancers while still curable. The Journal of Urology, 169(5), pp.1720-1723.

HUGOSSON, J., AUS, G., LILJA, H., LODDING, P. & PIHL, C. G. 2004. Results of a randomized, population-based study of biennial screening using serum prostate-specific antigen measurement to detect prostate carcinoma. Cancer: Interdisciplinary International Journal of the American Cancer Society, 100, 1397-1405.

HUMMEL, J. M., VAN ROSSUM, W., VERKERKE, G. J. & RAKHORST, G. 2000. The effects of team expert choice on group decision-making in collaborative new product development: a pilot study. Journal of Multicriteria Decision Analysis, 9, 90.

HUMMEL, J. M., BOOMKAMP, I. S., STEUTEN, L. M., VERKERKE, B. G. & IJZERMAN, M. J. 2012. Predicting the health economic performance of new non-fusion surgery in adolescent idiopathic scoliosis. Journal of Orthopaedic Research, 30, 1453-1458.

HUNINK, M.M., WEINSTEIN, M.C., WITTENBERG, E., DRUMMOND, M.F., PLISKIN, J.S., WONG, J.B. AND GLASZIOU, P.P., 2014. Decision making in health and medicine: integrating evidence and values. Cambridge University Press.

HUNTER, R. AND SHEARER, J (undated) Cost-consequences analysis - an underused method of economic evaluation, published by NIHR, UK. Available

from https://www.rds-london.nihr.ac.uk/resources/health-economics/cost-consequences-analysis-an-underused-method/ Accessed 8 July 2019

HWANG, T.J., SOKOLOV, E., FRANKLIN, J.M. AND KESSELHEIM, A.S., 2016. Comparison of rates of safety issues and reporting of trial outcomes for medical devices approved in the European Union and United States: cohort study. BMJ, 353.

IGLESIAS, C. P., THOMPSON, A., ROGOWSKI, W. H. & PAYNE, K. 2016. Reporting guidelines for the use of expert judgement in model-based economic evaluations. PharmacoEconomics, 34, 1161-1172.

IJZERMAN, M. J., KOFFIJBERG, H., FENWICK, E. & KRAHN, M. 2017. Emerging Use of Early Health Technology Assessment in Medical Product Development: A Scoping Review of the Literature. PharmacoEconomics, 1-14.

IJZERMAN, M. J. & STEUTEN, L. M. 2011. Early assessment of medical technologies to inform product development and market access. Applied Health Economics and Health Policy, 9, 331-347.

INAHTA. 2020. Welcome to INAHTA [online]. Available at www.inahta.org/ [Accessed 4 August 2020.]

INNOVATE UK. 2018. Home [online] Available at: https://www.gov.uk/government/organisations/innovate-uk [Accessed 31 July 2020].

INSTITUTE FOR CLINICAL AND ECONOMIC REVIEW. 2020. About. [online] Available at: https://icer-review.org/about/. [Accessed 31 July 2020].

INTERNATIONAL ERGONOMICS ASSOCIATION (undated). Human Factors/Ergonomics: Definition and applications. [online] Available from: https://iea.cc/what-is-ergonomics/ . [Accessed 31 July 2020]

ISD SCOTLAND. 2018a. Cancer | Publications | Data Tables | Health Topics | ISD Scotland. [online] Available at: https://www.isdscotland.org/Health-

Topics/Cancer/Publications/data-tables2017.asp?id=2276#2276 [Accessed 31 July 2020].

ISD SCOTLAND. 2018b. Cancer Waiting Times in NHS Scotland 1 April to 30 June 2018. [online] Available from https://www.isdscotland.org/Health-Topics/Waiting-Times/Publications/2018-09-25/2018-09-25-CWT-Report.pdf [Accessed 31 July 2020].

JONES, S. Coronavirus test kits withdrawn in Spain over poor accuracy rate. Guardian. March 27, 2020. Available at: https://www.theguardian.com/world/2020/mar/27/coronavirus-test-kits-withdrawn-spain-poor-accuracy-rate [Accessed 31 July 2020]

KAPLAN, A.V., BAIM, D.S., SMITH, J.J., FEIGAL, D.A., SIMONS, M., JEFFERYS, D., FOGARTY, T.J., KUNTZ, R.E. AND LEON, M.B., 2004. Medical device development: from prototype to regulatory approval. Circulation, 109(25), pp.3068-3072.

KENNIS RESEARCH. 2020. Receiver Operating Characteristic (ROC) Curves. [online]. Available at: https://kennis-research.shinyapps.io/ROC-Curves/ [Accessed 31 July 2020]

KIM, H., GOODALL, S. AND LIEW, D. 2020. The Potential for Early Health Economic Modelling in Health Technology Assessment and Reimbursement Decision-Making; Comment on “Problems and Promises of Health Technologies: The Role of Early Health Economic Modeling”. International Journal of Health Policy and Management.

KIP, M., STEUTEN, L. M., KOFFIJBERG, H., IJZERMAN, M. J. & KUSTERS, R. 2016. 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. Journal of Evaluation in Clinical Practice, 1-11.

KLUYTMANS, A., TUMMERS, M., VAN DER WILT, G.J. AND GRUTTERS, J., 2019. Early Assessment of Proof-of-Problem to Guide Health Innovation. Value in Health, 22(5), pp.601-606.

KOLOMINSKY-RABAS, P. L., DJANATLIEV, A., WAHLSTER, P., GANTNER-BÄR, M., HOFMANN, B., GERMAN, R., SEDLMAYR, M., REINHARDT, E., SCHÜTTLER, J. & KRIZA, C. 2015. Technology foresight for medical device development through hybrid simulation: The ProHTA Project. Technological Forecasting and Social Change, 97, 105-114.

KONING, R. 2012. Bacterial pathogen identification in blood samples with electronic nose technology in a clinical setting-An early medical technology assessment of the µDtect. Master's Thesis, University of Twente.

LATIMER, N., DIXON, S., MCDERMOTT, C., MCCARTHY, A., TINDALE, W., HERON, N. & SHAW, P. 2011. Modelling the cost effectiveness of a potential new neck collar for patients with motor neurone disease. Discussion Paper. HEDS Discussion Paper (11/10). Unpublished. Available at https://eprints.whiterose.ac.uk/43189/1/HEDS-DP_11-10.pdf [Accessed 3 February 2021]

LE, B. V., GRIFFIN, C. R., LOEB, S., CARVALHAL, G. F., KAN, D., BAUMANN, N. A. & CATALONA, W. J. 2010. [-2] Proenzyme prostate specific antigen is more accurate than total and free prostate specific antigen in differentiating prostate cancer from benign disease in a prospective prostate cancer screening study. The Journal of Urology, 183, 1355-1359.

LEDERMANN, J. A., RAJA, F. A., FOTOPOULOU, C., GONZALEZ-MARTIN, A., COLOMBO, N., SESSA, C. & ON BEHALF OF THE, E. G. W. G. 2013. Newly diagnosed and relapsed epithelial ovarian carcinoma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up†. Annals of Oncology, 24, vi24-vi32.

LEHOUX, P., WILLIAMS-JONES, B., MILLER, F., URBACH, D. AND TAILLIEZ, S., 2008. What leads to better health care innovation? Arguments for an integrated

policy-oriented research agenda. Journal of Health Services Research & Policy, 13(4), pp.251-254.

LEHOUX, P., MILLER, F.A., DAUDELIN, G. AND DENIS, J.L., 2017. Providing value to new health technology: the early contribution of entrepreneurs, investors, and regulatory agencies. International Journal of Health Policy and Management, 6(9), p.509.

LEHOUX, P. AND SILVA, H. 2020. Transforming Disciplinary Traditions; Comment on “Problems and Promises of Health Technologies: The Role of Early Health Economic Modeling”. International Journal of Health Policy and Management, 9(7), pp.309-311.

LEVIN, L. 2015. Early evaluation of new health technologies: The case for premarket studies that harmonize regulatory and coverage perspectives. International Journal of Technology Assessment in Health Care, 31, 207-209.

LINDEMANN, K., GIBBS, E., ÅVALL-LUNDQVIST, E., DEPONT CHRISTENSEN, R., WOIE, K., KALLING, M., AURANEN, A., GRENMAN, S., HOEGBERG, T. & ROSENBERG, P. 2017. Chemotherapy vs tamoxifen in platinum-resistant ovarian cancer: a phase III, randomised, multicentre trial (Ovaresist). British Journal of Cancer, 116, 455.

LOEB, S. & CATALONA, W. J. 2014. The Prostate Health Index: a new test for the detection of prostate cancer. Therapeutic Advances in Urology, 6, 74-77.

LOVE-KOH, J., 2019. How Useful Are Early Economic Models?: Comment on “Problems and Promises of Health Technologies: The Role of Early Health Economic Modelling”. International Journal of Health Policy and Management.

MADSEN, D.Ø., 2016. SWOT analysis: a management fashion perspective. International Journal of Business Research, 16(1), pp.39-56.

MANOVA, M., SAVOVA, A. T., VASILEVA, M. A., TEREZOVA, S., KAMUSHEVA, M., GREKOVA, D., PETKOVA, V. & PETROVA, G. I. 2018. Comparative Price Analysis

of Biological Products for Treatment of Rheumatoid Arthritis. Frontiers in Pharmacology, 9, 1070.

MARKIEWICZ, K., VAN TIL, J. A. & IJZERMAN, M. J. 2014. Medical devices early assessment methods: systematic literature review. International Journal of Technology Assessment in Health Care, 30, 137-146.

MARKIEWICZ, K., VAN TIL, J. A., STEUTEN, L. M. & IJZERMAN, M. J. 2016. Commercial viability of medical devices using Headroom and return on investment calculation. Technological Forecasting and Social Change, 112, 338-346.

MARKIEWICZ, K., VAN TIL, J. & IJZERMAN, M. 2017a. Early Assessment of Medical Devices in Development for Company Decision Making: An Exploration of Best Practices. Journal of Commercial Biotechnology, 23.

MARKIEWICZ, K. 2017b. Health technology assessment of medical devices during development. PhD Thesis, University of Twente.

MARS. 2020. Health. [online] Available at: https://www.marsdd.com/our-sectors/health/ [Accessed 31 July 2020].

MARTIN, J.L., NORRIS, B.J., MURPHY, E. AND CROWE, J.A., 2008. Medical device development: The challenge for ergonomics. Applied Ergonomics, 39(3), pp.271-283.

MAUSKOPF, J., SCHULMAN, K., BELL, L. & GLICK, H. 1996. A strategy for collecting pharmacoeconomic data during phase II/III clinical trials. Pharmacoeconomics, 9, 264-277.

MCCORMACK, R.T., ARMSTRONG, J. AND LEONARD, D., 2014. Co-development of genome-based therapeutics and companion diagnostics: insights from an Institute of Medicine roundtable. Journal of the American Medical Association, 311(14), pp.1395-1396.

MCATEER, H., COSH, E., FREEMAN, G., PANDIT, A., WOOD, P. & LILFORD, R. 2007. Cost-effectiveness analysis at the development phase of a potential health technology: examples based on tissue engineering of bladder and urethra. Journal of Tissue Engineering and Regenerative Medicine, 1, 343-349.

MCCLUGGAGE, W.G. My approach to and thoughts on the typing of ovarian carcinomas. Journal of Clinical Pathology. 2008;61(2):152-163. doi:10.1136/jcp.2007.049478.

MCMEEKIN, N., WU, O., GERMENI, E. AND BRIGGS, A., 2020. How methodological frameworks are being developed: evidence from a scoping review. BMC Medical Research Methodology, 20(1), pp.1-9.

MEDICAL SERVICES ADVISORY COMMITTEE. 2017. Technical Guidelines for preparing assessment reports for the Medical Services Advisory Committee – Service Type: Investigative (Version 3.0). [online] Available at: http://www.msac.gov.au/internet/msac/publishing.nsf/Content/0BD63667C984FEEACA25801000123AD8/$File/InvestigativeTechnicalGuidelines-December-2016-Version-3.0.pdf]. [Accessed 31 July 2020].

MEDICINES AND HEALTHCARE PRODUCTS REGULATORY AGENCY. Undated. Target Product Profile: Antibody tests to help determine if people have immunity to SARS-CoV-2. Version 2. [online] Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/883897/Target_Product_Profile_antibody_tests_to_help_determine_if_people_have_immunity_to_SARS-CoV-2_Version_2.pdf [Accessed 31 July 2020].

MEDICINES AND HEALTHCARE PRODUCTS REGULATORY AGENCY. 2020.Regulating Medical Devices in the UK. [online] Available at: https://www.gov.uk/guidance/regulating-medical-devices-in-the-uk [Accessed 3 February 2021].

MEDLINEPLUS. 2021. Acute vs.Chronic Conditions. [online] Available at https://medlineplus.gov/ency/imagepages/18126.htm [Accessed 3 February 2021].

MEDTECH EUROPE. 2010. FDA impact on US medical technology innovation: a survey of 200 firms. [online] Available at https://www.medtecheurope.org/wp-content/uploads/2015/07/01112010_FDA-impact-on-US-medical-technology-innovation_Backgrounder.pdf . [Accessed 31 July 2020].

MEDTECH EUROPE. 2019. The European Medical Technology Industry – in figures 2019. [online] Available at https://www.medtecheurope.org/wp-content/uploads/2019/04/The-European-Medical-Technology-Industry-in-figures-2019-2.pdf . [Accessed 31 July 2020].

MIKUDINA, B. & REDEKOP, K. 2013. Early medical technology assessments of medical devices and tests. Journal of Health Policy and Outcomes Research, 1, 26-37.

MILLER, P. 2005. Role of pharmacoeconomic analysis in R&D decision making. Pharmacoeconomics, 23, 1-12.

MIQUEL-CASES, A., 2016a. Early economic evaluation of technologies for emerging interventions to personalize breast cancer treatment. PhD thesis. Universite of Twente.

MIQUEL-CASES, A., RETÈL, V. P., VAN HARTEN, W. H. & STEUTEN, L. M. 2016b. Decisions on further research for predictive biomarkers of high-dose alkylating chemotherapy in triple-negative breast cancer: a value of information analysis. Value in Health, 19, 419-430.

MIQUEL-CASES, A., SCHOUTEN, P.C., STEUTEN, L. M., RETÈL, V. P., LINN, S.C. & VAN HARTEN, W. H. 2017. (Very) early technology assessment and translation of predictive biomarkers in breast cancer. Cancer Treatment Reviews 52, 117-127.

MORGAN, M.G., 2014. Use (and abuse) of expert elicitation in support of decision making for public policy. Proceedings of the National academy of Sciences, 111(20), pp.7176-7184.

MORRIS, D. E., OAKLEY, J. E. & CROWE, J. A. 2014. A web-based tool for eliciting probability distributions from experts. Environmental Modelling & Software, 52, 1-4.

MOTTET, N., BELLMUNT, J., BOLLA, M., BRIERS, E., CUMBERBATCH, M. G., DE SANTIS, M., FOSSATI, N., GROSS, T., HENRY, A. M. & JONIAU, S. 2017. EAU-ESTRO-SIOG guidelines on prostate cancer. Part 1: screening, diagnosis, and local treatment with curative intent. European Urology, 71, 618-629.

NATIONAL CANCER INSTITUTE. 2018. Cancer dictionary. Available from https://www.cancer.gov/publications/dictionaries/cancer-terms/search?contains=false&q=MSI [Accessed 31 July 2020]

NATIONAL HEALTH CARE INSTITUTE, NETHERLANDS. 2016. Guidelines for economic evaluations in health care. Available from https://english.zorginstituutnederland.nl/publications/reports/2016/06/16/guideline-for-economic-evaluations-in-healthcare [Accessed 31 July 2020].

NATHAVITHARANA, R.R., YOON, C., MACPHERSON, P., et al. Guidance for Studies Evaluating the Accuracy of Tuberculosis Triage Tests. J Infect Dis, Volume 220, Issue Supplement_3, 15 November 2019, Pages S116–S125, https://doi.org/10.1093/infdis/jiz243

NHS. 2018. Reference costs 2015 to 2016. [online] Available at: https://www.gov.uk/government/publications/nhs-reference-costs-2015-to-2016 [Accessed 31 July 2020].

NHS National Services Division. 2020. Molecular Pathology [online] Available at: https://www.nss.nhs.scot/specialist-healthcare/specialist-services/genetic-and-molecular-pathology-laboratories/ [Accessed 31 July 2020].

NICE. 2011. National Institute for Health and Care Excellence. Clinical Guideline 122 Ovarian Cancer: Recognition and Initial Management. Available from https://www.nice.org.uk/guidance/cg122 [Accessed 31 July 2020].

NICE. 2013. National Institute for Health and Care Excellence, Guide to the Methods of Technology Appraisal. Available at https:// https://www.nice.org.uk/process/pmg9/chapter/the-reference-case. [Accessed 31 July 2020].

NICE. 2014. National Institute for Health and Care Excellence. Clinical Guideline 131. Colorectal cancer: diagnosis and management [online] Available at: https://www.nice.org.uk/guidance/cg131 [Accessed 31 July 2020].

NICE. 2015. National Institute for Health and Care Excellence. Diagnostics guidance DG19. VivaScope 1500 and 3000 imaging systems for detecting skin cancer lesions. Available at: https://www.nice.org.uk/guidance/dg19/chapter/1-Recommendations [Accessed 31 July 2020].

NICE. 2016. National Institute for Health and Care Excellence, Clinical Guideline 181: Cardiovascular disease: risk assessment and reduction, including lipid modification. Available at: https://www.nice.org.uk/guidance/cg181/chapter/1-recommendations. [Accessed 31 July 2020].

NICE. 2018a. National Institute for Health and Care Excellence, Technology Assessment 242. Cetuximab, bevacizumab and panitumumab for the treatment of metastatic colorectal cancer after first-line chemotherapy: Cetuximab (monotherapy or combination chemotherapy), bevacizumab (in combination with non-oxaliplatin chemotherapy) and panitumumab(monotherapy) for the treatment of metastatic colorectal cancer after first-line chemotherapy. Available at: https://www.nice.org.uk/guidance/ta242 [Accessed 31 July 2020].

NICE. 2018b. National Institute for Health and Care Excellence. Diagnostics Assessment Programme. KRAS mutation testing of tumours in adults with metastatic colorectal cancer final scope. Available at: https://www.nice.org.uk/guidance/gid-dt14/documents/kras-mutation-testing-of-tumours-in-adults-with-metastatic-colorectal-cancer-final-scope2 [Accessed 31 July 2020].

NICE. 2018c. National Institute for Health and Care Excellence. Clinical Guideline 100. Rheumatoid arthritis in adults: management. NG100 [Online]. NICE. Available at: https://www.nice.org.uk/guidance/ng100 [Accessed 31 July 2020].

NICE. 2018d. National Institute for Health and Care Excellence. Diagnostics guidance DG34.Tumour profiling tests to guide adjuvant chemotherapy decisions in early breast cancer. Available at: https://www.nice.org.uk/guidance/dg34/chapter/1-Recommendations [Accessed 31 July 2020].

NICE. 2018e. Scientific advice. [online] Available at: https://www.nice.org.uk/about/what-we-do/life-sciences/scientific-advice [Accessed 31 July 2020].

NICE. 2019. National Institute for Health and Care Excellence, Clinical Guideline 131: Prostate cancer: diagnosis and management. Available at https:// https://www.nice.org.uk/guidance/ng131/chapter/Update-information [Accessed 31 July 2020].

NICE. 2020. Diagnostic assessment programme [online] Available at: https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/nice-diagnostics-guidance [Accessed 31 July 2020].

NICHOL, M., WU, J., AN, J., HUANG, J., DENHAM, D., FRENCHER, S. & JACOBSEN, S. 2011. Budget impact analysis of a new prostate cancer risk index for prostate cancer detection. Prostate Cancer and Prostatic Diseases, 14, 253.

NICHOL, M. B., WU, J., HUANG, J., DENHAM, D., FRENCHER, S. K. & JACOBSEN, S. J. 2012. Cost-effectiveness of Prostate Health Index for prostate cancer detection. BJU International, 110, 353-362.

NIHR. 2019. NIHR Medtech and In vitro diagnostics Co-operatives (MICs) [online] Available at: https://www.nihr.ac.uk/researchers/collaborations-services-and-support-for-your-research/find-services-or-support/access-methodology-advice.htm [Accessed 31 July 2020].

OAKLEY, J. & O'HAGAN, A. SHELF: the Sheffield Elicitation Framework (version 2.0). Sheffield: School of Mathematics and Statistics, University of Sheffield, 2010.

OECD. 2019. Health at a Glance 2019: OECD Indicators. [online] Available at: https://www.oecd-ilibrary.org/sites/4dd50c09-en/1/2/10/5/index.html?itemId=/content/publication/4dd50c09-en&_csp_=82587932df7c06a6a3f9dab95304095d&itemIGO=oecd&itemContentType=book#figure-d1e27486. [Accessed 31 July 2020].

OECD. 2020a. Health spending (indicator). [online] Available at: https://data.oecd.org/healthres/health-spending.htm. [Accessed 31 July 2020].

OECD. 2020b. Gross domestic product (GDP) (indicator). [online] Available at: https://data.oecd.org/gdp/gross-domestic-product-gdp.htm. [Accessed 31 July 2020]

O'HAGAN, A., STEVENS, J.W. AND CAMPBELL, M.J., 2005. Assurance in clinical trial design. Pharmaceutical Statistics: The Journal of Applied Statistics in the Pharmaceutical Industry, 4(3), pp.187-201.

OLIVER, S. R., REES, R. W., CLARKE-JONES, L., MILNE, R., OAKLEY, A. R., GABBAY, J., STEIN, K., BUCHANAN, P. & GYTE, G. 2008. A multidimensional conceptual framework for analysing public involvement in health services research. Health Expectations, 11, 72-84.

OOSTERHOFF, M., VAN DER MAAS, M.E. AND STEUTEN, L.M., 2016. A systematic review of health economic evaluations of diagnostic biomarkers. Applied Health Economics and Health Policy, 14(1), pp.51-65.

O'ROURKE, B., OORTWIJN, W. AND SCHULLER, T., 2020. The new definition of health technology assessment: a milestone in international collaboration. International Journal of Technology Assessment in Health Care, pp.1-4.

OSTERBERG, L. AND BLASCHKE, T., 2005. Adherence to medication. New England Journal of Medicine, 353(5), pp.487-497.

PAGE, P. 2006. Effectiveness of managing suspected pulmonary embolism using an algorithm combining clinical probability, D-dimer testing, and computed tomography. Journal of the Americal Medical Association. 295(2), 172-179.

PALEARI, L., GANDINI, S., PROVINCIALI, N., PUNTONI, M., COLOMBO, N. & DECENSI, A. 2017. Clinical benefit and risk of death with endocrine therapy in ovarian cancer: A comprehensive review and meta-analysis. Gynaecologic Oncology, 146, 504-513.

PARKINSON, D.R., MCCORMACK, R.T., KEATING, S.M., GUTMAN, S.I., HAMILTON, S.R., MANSFIELD, E.A., PIPER, M.A., DEVERKA, P., FRUEH, F.W., JESSUP, J.M. AND MCSHANE, L.M., 2014. Evidence of clinical utility: an unmet need in molecular diagnostics for patients with cancer. Clinical Cancer Research March 15 2014 (20) (6) 1428-1444.

PAYNE, K. AND ANNEMANS, L., 2013. Reflections on market access for personalized medicine: recommendations for Europe. Value in Health, 16(6), pp.S32-S38.

PBAC. 2016. Guidelines for preparing submissions to the Pharmaceutical Benefits Advisory Committee (PBAC). Version 5. September 2016. Available from https://pbac.pbs.gov.au/ [Accessed 31 July 2020].

PEEL, A., JENKS, M., CHOUDHURY, M., LOVETT, R., REJON-PARRILLA, J.C., SIMS, A. AND CRAIG, J., 2018. Use of Expert Judgement Across NICE Guidance-Making Programmes: A Review of Current Processes and Suitability of Existing Tools to Support the Use of Expert Elicitation. Applied Health Economics and Health Policy, 16(6), pp.819-836.

PETO, J. 2020. Covid-19 mass testing facilities could end the epidemic rapidly. BMJ. Mar 22;368.

PETTIT, S.D., JEROME, K.R., ROUQUIE, D. et al. 'All In': A Pragmatic Framework for COVID-19 Testing and Action on a Global Scale. EMBO Molecular Medicine. April 29, 2020.

PIETZSCH, J. B. & PATÉ-CORNELL, M. E. 2008. Early technology assessment of new medical devices. International Journal of Technology Assessment in Health Care, 24, 36-44.

PIRIE, M. 2015. How to win every argument: the use and abuse of logic. Bloomsbury, London, UK.

PORTER, D., VAN MELCKEBEKE, J., DALE, J., MESSOW, C. M., MCCONNACHIE, A., WALKER, A., MUNRO, R., MCLAREN, J., MCRORIE, E. & PACKHAM, J. 2016. Tumour necrosis factor inhibition versus rituximab for patients with rheumatoid arthritis who require biological treatment (ORBIT): an open-label, randomised controlled, non-inferiority, trial. The Lancet, 388, 239-247.

POSTMA, R., ROOBOL, M., SCHRÖDER, F. H. & VAN DER KWAST, T. H. 2004. Potentially advanced malignancies detected by screening for prostate carcinoma after an interval of 4 years. Cancer, 100, 968-975.

RETÈL, V. P., HUMMEL, J. M. & VAN HARTEN, W. H. 2008. Early phase Technology Assessment of nanotechnology in oncology. Tumori, 94, 284-290.

RETÈL, V. P., JOORE, M. A., LINN, S. C., RUTGERS, E. J. & VAN HARTEN, W. H. 2012. Scenario drafting to anticipate future developments in technology assessment. BMC Research Notes, 5, 442.

ROBERTS M, RUSSELL LB, PALTIEL AD, et al. Conceptualizing a model a report of the ISPOR-SMDM modeling good research practices task force–2. Medical Decision Making 2012;32:678–89

RODRIGUEZ, P.F., 2020. Predicting Whom to Test is More Important Than More Tests-Modeling the Impact of Testing on the Spread of COVID-19 Virus By True Positive Rate Estimation. Preprint available at:

https://www.medrxiv.org/content/10.1101/2020.04.01.20050393v1 [Accessed 31 July 2020].

ROGOWSKI, W., PAYNE, K., SCHNELL-INDERST, P., MANCA, A., ROCHAU, U., JAHN, B., ALAGOZ, O., LEIDL, R. AND SIEBERT, U., 2015. Concepts of 'personalization' in personalized medicine: implications for economic evaluation. Pharmacoeconomics, 33(1), pp.49-59.

RUBENSTEIN, J.H., VAKIL, N. AND INADOMI, J.M., 2005. The cost-effectiveness of biomarkers for predicting the development of oesophageal adenocarcinoma. Alimentary Pharmacology & Therapeutics, 22(2), pp.135-146.

RUSSELL, M.D., BUKHARI, M. AND GALLOWAY, J., 2018. The price of good health care. Rheumatology.

RYAN M, SCOTT DA, REEVES C, BATE A, VAN TEIJLINGEN ER, RUSSELL EM, et al. 2001. Eliciting public preferences for healthcare: a systematic review of techniques. Health Technology Assessment 5(5).

SACHS, R.E., 2015. Innovation law and policy: preserving the future of personalized medicine. UC Davis Law Review., 49, p.1881.

SAMMUT-BONNICI, T. AND GALEA, D., 2015. PEST analysis. Wiley Encyclopedia of management, pp.1-1.

SANTOS, R., URSU, O., GAULTON, A., BENTO, A.P., DONADI, R.S., BOLOGA, C.G., KARLSSON, A., AL-LAZIKANI, B., HERSEY, A., OPREA, T.I. AND OVERINGTON, J.P., 2017. A comprehensive map of molecular drug targets. Nature Reviews Drug Discovery, 16(1), p.19.

SCOTTISH MEDICINE CONSORTIUM 2018a Making a submission. [online] Available at: https://www.scottishmedicines.org.uk/making-a-submission/ [Accessed 22 Dec. 2018]

SCOTTISH MEDICINE CONSORTIUM. 2018b. Cetuximab (Erbitux). [online] Available at: https://www.scottishmedicines.org.uk/medicines-advice/cetuximab-erbitux-fullsubmission-101214/ [Accessed 31 July 2020].

SCHRÖDER, F. H., BANGMA, C. H. & ROOBOL, M. J. 2008. Is it necessary to detect all prostate cancers in men with serum PSA levels< 3.0 ng/ml? A comparison of biopsy results of PCPT and outcome-related information from ERSPC. European Urology, 53, 901-908.

SCULPHER, M., DRUMMOND, M. & BUXTON, M. 1997. The iterative use of economic evaluation as part of the process of health technology assessment. Journal of Health Services Research, 2, 26-30.

SECRETARY'S ADVISORY COMMITTEE ON GENETICS, HEALTH, AND SOCIETY. 2010. Gene Patents and Licensing Practices and Their Impact on Patient Access to Genetic Tests. [online] Available at: https://osp.od.nih.gov/wp-content/uploads/2013/11/SACGHS_patents_report_2010.pdf. [Accessed 31 July 2020].

SEER. 2014. Cancer Statistics Review 1975-2011. National Cancer Institute.

SIGN. 2011 Scottish Intercollegiate Guidelines Network. Publication 123. Management of early rheumatoid arthritis. Available from: https://www.sign.ac.uk/our-guidelines/management-of-early-rheumatoid-arthritis/ [Accessed 31 July 2020].

SIGN. 2017. Scottish Intercollegiate Guidelines Network. Publication 146 Cutaneous melanoma. Available from URL: https://www.sign.ac.uk/our-guidelines/cutaneous-melanoma/ [Accessed 31 July 2020].

SIGN. 2018. Scottish Intercollegiate Guidelines Network. Publication 126 Diagnosis and management of colorectal cancer. Available at: https://www.sign.ac.uk/our-guidelines/diagnosis-and-management-of-colorectal-cancer/ [Accessed 31 July 2020].

SINGH, P. 2010. Enabling late-stage translation of regenerative medicine-based products. PhD Thesis, Loughborough University.

SPENCER, L. & RITCHIE, J. 2002. Qualitative data analysis for applied policy research. Analyzing Qualitative Data. Routledge.

STATISTA. Undated, a. Prescription drugs market shares by top companies globally. Available at: https://www.statista.com/statistics/309425/prescription-drugs-market-shares-by-top-companies-globally/. [Accessed 31 July 2020].

STATISTA. Undated, b. Revenue of global top medical technology companies. Available at: https://www.statista.com/statistics/281544/revenue-of-global-top-medical-technology-companies/. [Accessed 31 July 2020].

STATISTA. Undated, c. Global medtech market share of top 20 companies. Available at: https://www.statista.com/statistics/329035/global-medtech-market-share-of-top-20-companies/. [Accessed 31 July 2020].

STEUTEN, L., VALLEJO-TORRES, L., BASTIDE, P. & BUXTON, M. 2009. Analysing uncertainty around costs of innovative medical technologies: the case of fibrin sealant (QUIXIL®) for total knee replacement. Health Policy, 89, 46-57.

STEUTEN, L., VAN DE WETERING, G., GROOTHUIS-OUDSHOORN, K. & RETÈL, V. 2013. A systematic and critical review of the evolving methods and applications of value of information in academia and practice. Pharmacoeconomics, 31, 25-48.

STEUTEN, L. M. 2016. Multi-Dimensional Impact of the Public–Private Center for Translational Molecular Medicine (CTMM) in the Netherlands: Understanding New 21st Century Institutional Designs to Support Innovation-in-Society. Omics: a Journal of Integrative Biology, 20, 265-273.

STEVENS, A., MILNE, R. AND BURLS, A. 2003. Health technology assessment: history and demand. Journal of Public Health, 25(2), pp.98-101.

STEVENSON, M., ARCHER, R., TOSH, J., SIMPSON, E., EVERSON-HOCK, E., STEVENS, J., HERNANDEZ-ALAVA, M., PAISLEY, S., DICKINSON, K. & SCOTT, D. 2016. Adalimumab, etanercept, infliximab, certolizumab pegol, golimumab, tocilizumab and abatacept for the treatment of rheumatoid arthritis not previously treated with disease-modifying antirheumatic drugs and after the failure of conventional disease-modifying antirheumatic drugs only: systematic review and economic evaluation. Health Technology Assessment, 20, 1-610.

STEYERBERG, E.W., VICKERS, A.J., COOK, N.R., GERDS, T., GONEN, M., OBUCHOWSKI, N., PENCINA, M.J. AND KATTAN, M.W., 2010. Assessing the performance of prediction models: a framework for some traditional and novel measures. Epidemiology (Cambridge, Mass.), 21(1), p.128.

STEYERBERG, E.W., VEDDER, M.M., LEENING, M.J., POSTMUS, D., D'AGOSTINO SR, R.B., VAN CALSTER, B. AND PENCINA, M.J., 2015. Graphical assessment of incremental value of novel markers in prediction models: from statistical to decision analytical perspectives. Biometrical Journal, 57(4), pp.556-570.

TEOH, J.Y.C., LEUNG, C.H., WANG, M.H., CHIU, P.K.F., YEE, C.H., NG, C.F. AND WONG, M.C.S., 2020. The cost-effectiveness of prostate health index for prostate cancer detection in Chinese men. Prostate Cancer and Prostatic Diseases, pp.1-7.

THE VALUE INITIATIVE. 2020. Open Source Value Project. [online] Available at: https://www.thevalueinitiative.org/open-source-value-project/ [Accessed 31 July 2020]

THOMPSON, A. J., NEWMAN, W. G., ELLIOTT, R. A., ROBERTS, S. A., TRICKER, K. & PAYNE, K. 2014. The cost-effectiveness of a pharmacogenetic test: a trial-based evaluation of TPMT genotyping for azathioprine. Value in Health, 17, 22-33.

TOWSE, A., OSSA, D., VEENSTRA, D., CARLSON, J. & GARRISON, L. 2013. Understanding the economic value of molecular diagnostic tests: case studies and lessons learned. Journal of Personalized Medicine, 3, 288-305.

TRUSHEIM, M.R., BERNDT, E.R. AND DOUGLAS, F.L., 2007. Stratified medicine: strategic and economic implications of combining drugs and clinical biomarkers. Nature Reviews Drug Discovery, 6(4), p.287.

USACH, I., BLANSIT, K., CHEN, L.M., et al. Survival differences in women with serous tubal, ovarian, peritoneal, and uterine carcinomas. American Journal of Obsteterics and Gynecology. 2015;212(2):188.e1-6. doi:10.1016/j.ajog.2014.08.016.

U.S. Food and Drug Administration. 2017. Classification of Products as Drugs and Devices and Additional Product Classification Issues Guidance for Industry and FDA Staff. [online]. Available at https://www.fda.gov/regulatory-information/search-fda-guidance-documents/classification-products-drugs-and-devices-and-additional-product-classification-issues. [Accessed 4 August 2020].

U.S. Food and Drug Administration. 2018. Laboratory Developed Tests. [online] Available at: https://www.fda.gov/medical-devices/vitro-diagnostics/laboratory-developed-tests. [Accessed 4 August 2020].

U.S. Food and Drug Administration. 2020a. Coronavirus (COVID-19) April 20, 2020. Update: FDA Authorizes First Test for Patient At-Home Sample Collection [online] Available at: https://www.fda.gov/news-events/press-announcements/coronavirus-covid-19-update-fda-authorizes-first-test-patient-home-sample-collection [Accessed 4 August 2020]

U.S. Food and Drug Administration. 2020b. Emergency Use Authorisations [online] Available at: https://www.fda.gov/medical-devices/emergency-situations-medical-devices/emergency-use-authorizations#covid19ivd [Accessed 4 August 2020]

VALLEJO-TORRES, L., STEUTEN, L., PARKINSON, B., GIRLING, A. J. & BUXTON, M. J. 2011. Integrating health economics into the product development cycle: a case study of absorbable pins for treating hallux valgus. Medical Decision Making, 31, 596-610.

VALLEJO-TORRES, L., STEUTEN, L. M., BUXTON, M. J., GIRLING, A. J., LILFORD, R. J. & YOUNG, T. 2008. Integrating health economics modeling in the product development cycle of medical devices: a Bayesian approach. International Journal of Technology Assessment in Health Care, 24, 459-464.

VAN DER CRUIJSEN-KOETER, I. W., VAN DER KWAST, T. H. & SCHRØDER, F. H. 2003. Interval carcinomas in the European randomized study of screening for prostate cancer (ERSPC)–Rotterdam. Journal of the National Cancer Institute, 95, 1462-1466.

VAN NIMWEGEN, K. J., LILFORD, R. J., VAN DER WILT, G. J. & GRUTTERS, J. P. 2017. HEADROOM BEYOND THE QUALITY-ADJUSTED LIFE-YEAR: THE CASE OF COMPLEX PEDIATRIC NEUROLOGY. International Journal of Technology Assessment in Health Care, 1-6.

VAN NORMAN, G.A., 2016a. Drugs, devices, and the FDA: Part 1: an overview of approval processes for drugs. JACC: Basic to Translational Science, 1(3), pp.170-179.

VAN NORMAN, G.A., 2016b. Drugs, devices, and the FDA: part 2: an overview of approval processes: FDA approval of medical devices. JACC: Basic to Translational Science, 1(4), pp.277-287.

VILSBØLL, A.W., MOURITSEN, J.M., JENSEN, L.P., BØDKER, N., HOLST, A.W., PENNISI, C.P. AND EHLERS, L., 2018. Cell-based therapy for the treatment of female stress urinary incontinence: an early cost–effectiveness analysis. Regenerative Medicine, 13(3), pp.321-330.

WANG, Y., ZHANG S., LI F., ZHOU, Y., ZHANG, W., WANG, Z., ZHANG, R., ZHU, J., REN, Y., TAN, Y., QIN, C., LI, Y., LI, X., CHEN Y., ZHU, F. Therapeutic target database 2020: enriched resource for facilitating research and early development of targeted therapeutics, Nucleic Acids Research, Volume 48, Issue D1, 08 January 2020, Pages D1031–D1041.

WEDDERBURN, C.J., WARNER, P., GRAHAM, B., DUNCAN, W.C., CRITCHLEY, H.O.D. AND HORNE, A.W., 2009. Economic evaluation of diagnosing and excluding ectopic pregnancy. Human Reproduction, 25(2), pp.328-333.

WEISER, M. K. 2013. A photoacoustic instrument for diagnosis and monitoring of rheumatoid arthritis-A case study in the field of user involvement. Master's Thesis, Universite of Twente.

WESTWOOD, M., LANG, S,, ARMSTRONG, N. et al. Faecal immunochemical tests (FIT) can help to rule out colorectal cancer in patients presenting in primary care with lower abdominal symptoms: a systematic review conducted to inform new NICE DG30 diagnostic guidance. BMC Medicine 15, 189 (2017). https://doi.org/10.1186/s12916-017-0944-z

WISSING, T.B., 2012. The Bioartificial Pancreas" A Clinical Case Scenario Analysis to assess and support the development of a device to improve type I Diabetes care". Master's Thesis, Universite of Twente.

WOLTERS, T., VAN DER KWAST, T. H., VISSERS, C. J., BANGMA, C. H., ROOBOL, M., SCHRÖDER, F. H. & VAN LEENDERS, G. J. 2010. False-negative prostate needle biopsies: frequency, histopathologic features, and follow-up. The American Journal of Surgical Pathology, 34, 35-43.

WORLD HEALTH ORGANISATION. 2017. Global Model Regulatory Framework for Medical Devices including in vitro diagnostic medical devices. [online] Available at: https://www.who.int/medical_devices/publications/global_model_regulatory_framework_meddev/en/ [Accessed 31 July 2020].

WORLD HEALTH ORGANISATION. 2020. Health Technology Assessment. HTA Definitions [online] Available at: https://www.who.int/health-technology-assessment/about/Defining/en/ [Accessed 31 July 2020].

WOSCAN. 2018. West of Scotland Cancer Network. [online] Available at: https://www.woscan.scot.nhs.uk/ [Accessed 31 July 2020].

YOCK, P.G., ZENIOS, S., MAKOWER, J., BRINTON, T.J., KUMAR, U.N., WATKINS, F.J., DENEND, L., KRUMMEL, T.M. AND KURIHARA, C.Q., 2015. Biodesign: the process of innovating medical technologies. Cambridge University Press.

YORK HEALTH ECONOMICS CONSORTIUM. 2016. Glossary: Discrete Choice Experiment (DCE) [online]. Available at : https://yhec.co.uk/glossary/discrete-choice-experiment-dce/ [Accessed 31 July 2020].

YUAN, Z.-X., WANG, X.-Y., QIN, Q.-Y., CHEN, D.-F., ZHONG, Q.-H., WANG, L. & WANG, J.-P. 2013. The prognostic role of BRAF mutation in metastatic colorectal cancer receiving anti-EGFR monoclonal antibodies: a meta-analysis. PLoS One, 8, e65995.

ZHOU, X.H., MCCLISH, D.K. AND OBUCHOWSKI, N.A. 2009. Statistical methods in diagnostic medicine (Vol. 569). John Wiley & Sons.

ZWEIG, M.H. AND CAMPBELL, G., 1993. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clinical Chemistry, 39(4), pp.561-577.

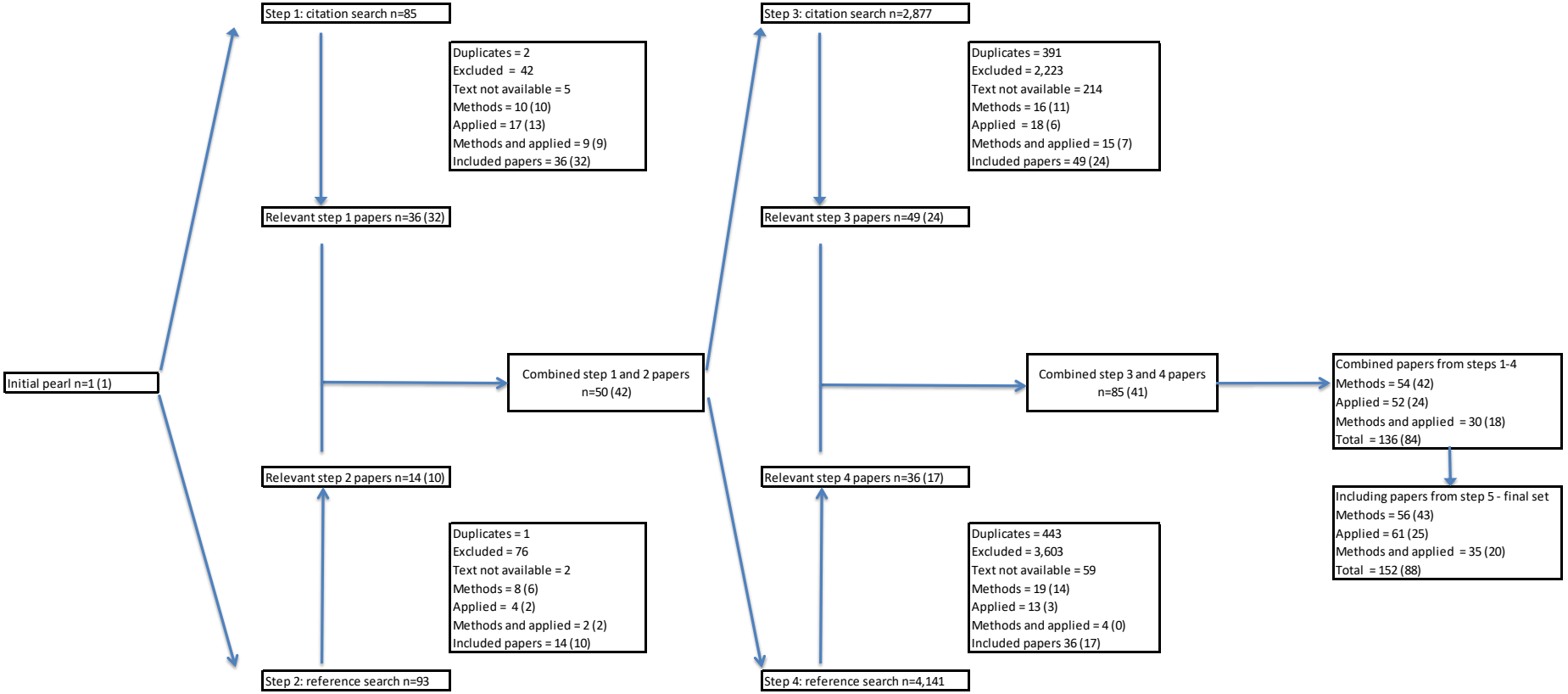
List of appendices

Appendices

Appendix 1 – Flowchart of literature review

Flowchart of literature review

Total studies with number of development-focused studies in brackets



Notes

Step 1 is a search of the citations on Google Scholar of the initial pearl article (Ijzerman and Steuten, 2011)

Step 2 is a search of the references of the initial pearl (Ijzerman and Steuten, 2011)

Step 3 is a search of the citations of all the identified articles from steps 1 and 2.

Step 4 is a search of all the references of all the identified articles from steps 1 and 2.

Step 5 is a search of citations of all articles identified as having explicit or implicit features of development-focused HTA (listed in Appendices) in February 2019

Appendix 2 - Papers extracted in literature review

Methods papers – HTA to inform developers

ANNEMANS, L., GENESTÉ, B. & JOLAIN, B. 2000. Early modelling for assessing health and economic outcomes of drug therapy. Value in Health, 3, 427-434.

BARTELMES, M., NEUMANN, U., LÜHMANN, D., SCHÖNERMARK, M. P. & HAGEN, A. 2009. Methods for assessment of innovative medical technologies during early stages of development. GMS health technology assessment, 5.

BELL, D., KASHEFI, A., SALEH, N., TURCHI, T. & YOUNG, T. A data-driven agent based simulation platform for early health economics device evaluation. Proceedings of the Agent-Directed Simulation Symposium, 2016. Society for Computer Simulation International, 5.

BORSCI, S., BUCKLE, P., UCHEGBU, I., NI, M., WALNE, S. & HANNA, G. B. 2017. Integrating human factors and health economics to inform the design of medical device: a conceptual framework. EMBEC & NBC 2017. Springer.

BOYD, K. A., FENWICK, E. & BRIGGS, A. 2010. Using an iterative approach to economic evaluation in the drug development process. Drug Development Research, 71, 470-477.

CLEMENS, K., GARRISON, L. P., JONES, A. & MACDONALD, F. 1993. Strategic use of pharmacoeconomic research in early drug development and global pricing. Pharmacoeconomics, 4, 315-322.

COSH, E., GIRLING, A., LILFORD, R., MCATEER, H. & YOUNG, T. 2007. Investing in new medical technologies: a decision framework. Journal of Commercial Biotechnology, 13, 263-271.

CRAVEN, M. P., MORGAN, S. P., CROWE, J. A. & LU, B. 2009. Deploying a spreadsheet tool for early economic value assessment of medical device innovations with healthcare decision makers. Journal of Management & Marketing in Healthcare, 2, 278-292.

FASTERHOLDT, I., KRAHN, M., KIDHOLM, K., YDERSTRÆDE, K. B. & PEDERSEN, K. M. 2017. Review of early assessment models of innovative medical technologies. Health Policy, 121, 870-879.

FREMPONG, S. N., SUTTON, A. J., DAVENPORT, C. & BARTON, P. 2018. Economic evaluation of medical tests at the early phases of development: a systematic review of empirical studies. Expert review of pharmacoeconomics & outcomes research, 18, 13-23.

GIRLING, A., LILFORD, R., COLE, A. & YOUNG, T. 2015. Headroom approach to device development: current and future directions. International journal of technology assessment in health care, 31, 331-338.

GIRLING, A., YOUNG, T., BROWN, C. & LILFORD, R. 2010. Early-stage valuation of medical devices: the role of developmental uncertainty. Value in Health, 13, 585-591.

GIRLING, A. J., LILFORD, R. J. & YOUNG, T. P. 2012. Pricing of medical devices under coverage uncertainty—a modelling approach. Health economics, 21, 1502-1507.

GOSLING, J. P. 2014. Methods for eliciting expert opinion to inform health technology assessment. Vignette Commissioned by the MRC Methodology Advisory Group. Medical Research Council (MRC) and National Institute for Health Research (NIHR).

GRABOWSKI, H. 1997. The effect of pharmacoeconomics on company research and development decisions. Pharmacoeconomics, 11, 389-397.

HARTZ, S. & JOHN, J. 2008. Contribution of economic evaluation to decision making in early phases of product development: a methodological and empirical review. International Journal of Technology Assessment in Health Care, 24, 465-472.

HEMELS, M., WOLDEN, M. & EINARSON, T. R. 2009. A Matrix to Determine Health Economic Viability Throughout Product Development: A Pharmaceutical Industry Perspective. Drug Information Journal, 43, 749-756.

HUGHES, D. A. & WALLEY, T. 2001. Economic evaluations during early (phase II) drug development. Pharmacoeconomics, 19, 1069-1077.

HUMMEL, M. J., VAN ROSSUM, W., VERKERKE, G. J. & RAKHORST, G. 2000. Assessing medical technologies in development: A new paradigm of medical technology assessment. International Journal of Technology Assessment in Health Care, 16, 1214-1219.

IJZERMAN, M. J., KOFFIJBERG, H., FENWICK, E. & KRAHN, M. 2017. Emerging Use of Early Health Technology Assessment in Medical Product Development: A Scoping Review of the Literature. PharmacoEconomics, 1-14.

IJZERMAN, M. J. & STEUTEN, L. M. 2011. Early assessment of medical technologies to inform product development and market access. Applied health economics and health policy, 9, 331-347.

LANGLEY, P. C. 2004. Focusing pharmacoeconomic activities: reimbursement or the drug life cycle? Current medical research and opinion, 20, 181-188.

LEE, J., TOLLEFSON, E., DALY, M. & KIELB, E. 2013. A generalized health economic and outcomes research model for the evaluation of companion diagnostics and targeted therapies. Expert review of pharmacoeconomics & outcomes research, 13, 361-370.

LEVIN, L. 2015. Early evaluation of new health technologies: The case for premarket studies that harmonize regulatory and coverage perspectives. International journal of technology assessment in health care, 31, 207-209.

LIBERATORE, M. J. & NYDICK, R. L. 2008. The analytic hierarchy process in medical and health care decision making: A literature review. European Journal of Operational Research, 189, 194-207.

LU, B., MARTIN, J. L., CRAVEN, M. P. & MORGAN, S. P. 2011. Can health economics aid decision making in healthcare innovation in academia? International Journal of Technology Transfer and Commercialisation, 11, 70-87.

MARKIEWICZ, K., VAN TIL, J. & IJZERMAN, M. 2017. Early Assessment of Medical Devices in Development for Company Decision Making: An Exploration of Best Practices. Journal of Commercial Biotechnology, 23.

MARKIEWICZ, K., VAN TIL, J. A. & IJZERMAN, M. J. 2014. Medical devices early assessment methods: systematic literature review. International journal of technology assessment in health care, 30, 137-146.

MIKUDINA, B. & REDEKOP, K. 2013. Early medical technology assessments of medical devices and tests. Journal of Health Policy and Outcomes Research, 1, 26-37.

MILLER, P. 2005. Role of pharmacoeconomic analysis in R&D decision making. Pharmacoeconomics, 23, 1-12.

MIQUEL CASES, A. 2016. Early economic evaluation of technologies for emerging interventions to personalize breast cancer treatment, Universiteit Twente.

MIQUEL-CASES, A., SCHOUTEN, P. C., STEUTEN, L. M., RETÈL, V. P., LINN, S. C. & VAN HARTEN, W. H. 2017. (Very) Early technology assessment and translation of predictive biomarkers in breast cancer. Cancer treatment reviews, 52, 117-127.

MORRIS, D. E., OAKLEY, J. E. & CROWE, J. A. 2014. A web-based tool for eliciting probability distributions from experts. Environmental Modelling & Software, 52, 1-4.

POSTMA, T. J., BROEKHUIZEN, T. L. & VAN DEN BOSCH, F. 2012. The contribution of scenario analysis to the front-end of new product development. Futures, 44, 642-654.

RETÈL, V. P., HUMMEL, M. J. & VAN HARTEN, W. H. 2009. Review on early technology assessments of nanotechnologies in oncology. Molecular oncology, 3, 394-401.

SCHOT, J. & RIP, A. 1997. The past and future of constructive technology assessment. Technological forecasting and social change, 54, 251-268.

STEUTEN, L., VAN DE WETERING, G., GROOTHUIS-OUDSHOORN, K. & RETÈL, V. 2013. A systematic and critical review of the evolving methods and applications of value of information in academia and practice. Pharmacoeconomics, 31, 25-48.

STEUTEN, L. M. & RAMSEY, S. D. 2014. Improving early cycle economic evaluation of diagnostic technologies. Expert review of pharmacoeconomics & outcomes research, 14, 491-498.

STEYERBERG, E. W., VEDDER, M. M., LEENING, M. J., POSTMUS, D., D'AGOSTINO, R. B., VAN CALSTER, B. & PENCINA, M. J. 2015. Graphical assessment of incremental value of novel markers in prediction models: From statistical to decision analytical perspectives. Biometrical Journal, 57, 556-570.

STEYERBERG, E. W., VICKERS, A. J., COOK, N. R., GERDS, T., GONEN, M., OBUCHOWSKI, N., PENCINA, M. J. & KATTAN, M. W. 2010. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology, 21, 128-38.

TUFFAHA, H. W., GORDON, L. G. & SCUFFHAM, P. A. 2013. Value of information analysis in oncology: the value of evidence and evidence of value. Journal of oncology practice, 10, e55-e62.

VALLEJO-TORRES, L., STEUTEN, L. M., BUXTON, M. J., GIRLING, A. J., LILFORD, R. J. & YOUNG, T. 2008. Integrating health economics modeling in the product development cycle of medical devices: a Bayesian approach. International journal of technology assessment in health care, 24, 459-464.

WAHLSTER, P., GOETGHEBEUR, M., KRIZA, C., NIEDERLÄNDER, C. & KOLOMINSKY-RABAS, P. 2015. Balancing costs and benefits at different stages of medical innovation: a systematic review of Multi-criteria decision analysis (MCDA). BMC health services research, 15, 262.

Applied papers – HTA to inform developers

BRAZ, G. E. R. 2013. Modeling key uncertainties in technology development: The case of Twente Photoacoustic Mammoscope (PAM).

BUISMAN, L. R., LUIME, J. J., OPPE, M., HAZES, J. M. & RUTTEN-VAN MÖLKEN, M. P. 2016a. A five-year model to assess the early cost-effectiveness of new diagnostic tests in the early diagnosis of rheumatoid arthritis. Arthritis research & therapy, 18, 135.

CAO, Q., POSTMUS, D., HILLEGE, H. L. & BUSKENS, E. 2013. Probability elicitation to inform early health economic evaluations of new medical technologies: a case study in heart failure disease management. Value in Health, 16, 529-535.

CRAVEN, M. P. & MORGAN, S. P. 2011. Early Stage Economic Evaluation with a Small Medical Device Start-up Company using a Markov Model. Transactions of the ASME-W-Journal of Medical Devices, 5, 027516.

GAULTNEY, J. G., SANHUEZA, E., JANSSEN, J. J., REDEKOP, W. K. & UYL-DE GROOT, C. A. 2011. Application of cost–effectiveness analysis to demonstrate the potential value of companion diagnostics in chronic myeloid leukemia. Pharmacogenomics, 12, 411-421.

GROOTHUIS-OUDSHOORN, C. G., FERMONT, J. M., VAN TIL, J. A. & IJZERMAN, M. J. 2014. Public stated preferences and predicted uptake for genome-based colorectal cancer screening. BMC medical informatics and decision making, 14, 18.

HILGERINK, M. P., HUMMEL, M. J., MANOHAR, S., VAARTJES, S. R. & IJZERMAN, M. J. 2011. Assessment of the added value of the Twente Photoacoustic Mammoscope in breast cancer diagnosis. Medical devices (Auckland, NZ), 4, 107.

HJELMGREN, J., GHATNEKAR, O., REIMER, J., GRABOWSKI, M., LINDVALL, O., PERSSON, U. & HAGELL, P. 2006. Estimating the value of novel interventions for

Parkinson's disease: an early decision-making model with application to dopamine cell replacement. Parkinsonism & related disorders, 12, 443-452.

HUMMEL, J. M., BOOMKAMP, I. S., STEUTEN, L. M., VERKERKE, B. G. & IJZERMAN, M. J. 2012. Predicting the health economic performance of new non-fusion surgery in adolescent idiopathic scoliosis. Journal of orthopaedic research, 30, 1453-1458.

HUYGENS, S. A., RUTTEN-VAN MÖLKEN, M. P., BEKKERS, J. A., BOGERS, A. J., BOUTEN, C. V., CHAMULEAU, S. A., DE JAEGERE, P. P., KAPPETEIN, A. P., KLUIN, J. & VAN MIEGHEM, N. M. 2016. Conceptual model for early health technology assessment of current and novel heart valve interventions. Open heart, 3, e000500.

KENTER, W. 2015. Exploring Markov modeling approaches for the health economic assessment of circulating tumor cells in the management of metastatic prostate cancer patients.

KHOUDIGIAN-SINANI, S., BLACKHOUSE, G., LEVINE, M., THABANE, L. & O'REILLY, D. 2017. The premarket assessment of the cost-effectiveness of a predictive technology "Straticyte™" for the early detection of oral cancer: a decision analytic model. Health Economics Review, 7, 35.

KIP, M., STEUTEN, L. M., KOFFIJBERG, H., IJZERMAN, M. J. & KUSTERS, R. 2016. 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. Journal of evaluation in clinical practice, 1-11.

KNUTTEL, F. M., HUIJSSE, S. E., FEENSTRA, T. L., MOONEN, C. T., VAN DEN BOSCH, M. A., BUSKENS, E., GREUTER, M. J. & DE BOCK, G. H. 2017. Early health technology assessment of magnetic resonance-guided high intensity focused ultrasound ablation for the treatment of early-stage breast cancer. Journal of therapeutic ultrasound, 5, 23.

KOERBER, F., ROLAUFFS, B. & ROGOWSKI, W. 2013. Early evaluation and value-based pricing of regenerative medicine technologies. Regenerative medicine, 8, 747-758.

KONING, R. 2012. Bacterial pathogen identification in blood samples with electronic nose technology in a clinical setting-An early medical technology assessment of the µDtect. University of Twente.

LATIMER, N., DIXON, S., MCDERMOTT, C., MCCARTHY, A., TINDALE, W., HERON, N. & SHAW, P. 2011. Modelling the cost effectiveness of a potential new neck collar for patients with motor neurone disease.

LIU, S. X., LEE, M. C., ATAKHORRAMI, M., TATOUSEK, J., MCCORMACK, M., YUNG, R., HART, N. & WHITE, D. P. 2013. Economic assessment of home-based COPD management programs. COPD: Journal of Chronic Obstructive Pulmonary Disease, 10, 640-649.

LUIME, J. J., BUISMAN, L. R., OPPE, M., HAZES, J. M. & RUTTEN-VAN MÖLKEN, M. P. 2016. Cost-Effectiveness Model for Evaluating New Diagnostic Tests in the Evaluation of Patients With Inflammatory Arthritis at Risk of Having Rheumatoid Arthritis. Arthritis care & research, 68, 927-935.

MCATEER, H., COSH, E., FREEMAN, G., PANDIT, A., WOOD, P. & LILFORD, R. 2007. Cost-effectiveness analysis at the development phase of a potential health technology: examples based on tissue engineering of bladder and urethra. Journal of tissue engineering and regenerative medicine, 1, 343-349.

REJON-PARRILLA, J. C., NUIJTEN, M., REDEKOP, W. K. & GAULTNEY, J. G. 2014. Economic evaluation of the use of a pharmacogenetic diagnostic test in schizophrenia. Health Policy and Technology, 3, 314-324.

VAN DE WETERING, G., STEUTEN, L. M., VON BIRGELEN, C., ADANG, E. M. & IJZERMAN, M. J. 2012. Early Bayesian modeling of a potassium lab-on-a-chip for monitoring of heart failure patients at increased risk of hyperkalaemia. Technological forecasting and social change, 79, 1268-1279.

VILSBØLL, A.W., MOURITSEN, J.M., JENSEN, L.P., BØDKER, N., HOLST, A.W., PENNISI, C.P. AND EHLERS, L., 2018. Cell-based therapy for the treatment of female stress urinary incontinence: an early cost–effectiveness analysis. Regenerative medicine, 13(3), pp.321-330.

WEISER, M. K. 2013. A photoacoustic instrument for diagnosis and monitoring of rheumatoid arthritis-A case study in the field of user involvement.

WISSING, T. 2012. The Bioartificial Pancreas" A Clinical Case Scenario Analysis to assess and support the development of a device to improve type I Diabetes care".

Combined methods and applied papers – HTA to inform developers

BRANDES, A., SINNER, M. F., KÄÄB, S. & ROGOWSKI, W. H. 2015. Early decision-analytic modeling–a case study on vascular closure devices. BMC health services research, 15, 486.

BRETELER, M. 2012. Scenario Analysis and Real options modeling of home brain monitoring in epilepsy patients. University of Twente.

BUISMAN, L. R., RUTTEN-VAN MÖLKEN, M. P., POSTMUS, D., LUIME, J. J., UYL-DE GROOT, C. A. & REDEKOP, W. K. 2016(b). The early bird catches the worm: early cost-effectiveness analysis of new medical tests. International journal of technology assessment in health care, 32, 46-53.

CARR, S. 2013. Simulation-based Analysis of Cell-based Regenerative Medicine Production: Focus on Efficiency and Cost-effectiveness, North Carolina State University.

CHAPMAN, A.M., 2013. The use of early economic evaluation to inform medical device decisions: an evaluation of the Headroom method (Doctoral dissertation, University of Birmingham).

DAVEY, S. M., BRENNAN, M., MEENAN, B. J., MCADAM, R., GIRLING, A., CHAPMAN, A. & LILFORD, R. L. 2011. A framework to manage the early value proposition of emerging healthcare technologies. Irish Journal of Management, 31, 59.

DE GRAAF, G., POSTMUS, D. & BUSKENS, E. 2015. Using multicriteria decision analysis to support research priority setting in biomedical translational research projects. BioMed research international, 2015.

DE GRAAF, G., POSTMUS, D., WESTERINK, J. & BUSKENS, E. 2018. The early economic evaluation of novel biomarkers to accelerate their translation into clinical applications. Cost Effectiveness and Resource Allocation, 16, 23.

GANTNER-BÄR, M., DJANATLIEV, A., PROKOSCH, H.-U. & SEDLMAYR, M. 2011. Conceptual modeling for Prospective Health Technology Assessment. Studies in health technology and informatics, 180, 33-37.

HAAKMA, W., STEUTEN, L. M., BOJKE, L. & IJZERMAN, M. J. 2014. Belief elicitation to populate health economic models of medical diagnostic devices in development. Applied health economics and health policy, 12, 327-334.

KLUYTMANS, A., TUMMERS, M., VAN DER WILT, G. J. & GRUTTERS, J. 2019. Early Assessment of Proof-of-Problem to Guide Health Innovation. Value in Health.

KOLOMINSKY-RABAS, P. L., DJANATLIEV, A., WAHLSTER, P., GANTNER-BÄR, M., HOFMANN, B., GERMAN, R., SEDLMAYR, M., REINHARDT, E., SCHÜTTLER, J. & KRIZA, C. 2015. Technology foresight for medical device development through hybrid simulation: The ProHTA Project. Technological Forecasting and Social Change, 97, 105-114.

MARKIEWICZ, K., VAN TIL, J. A., STEUTEN, L. M. & IJZERMAN, M. J. 2016. Commercial viability of medical devices using Headroom and return on investment calculation. Technological forecasting and social change, 112, 338-346.

MARKIEWICZ, K., VAN TIL, J. & IJZERMAN, M. 2017. Early Assessment of Medical Devices in Development for Company Decision Making: An Exploration of Best Practices. Journal of Commercial Biotechnology, 23.

PIETZSCH, J. B. & PATÉ-CORNELL, M. E. 2008. Early technology assessment of new medical devices. International Journal of Technology Assessment in Health Care, 24, 36-44.

POSTMUS, D., GRAAF, G., HILLEGE, H. L., STEYERBERG, E. W. & BUSKENS, E. 2012. A method for the early health technology assessment of novel biomarker measurement in primary prevention programs. Statistics in medicine, 31, 2733-2744.

SINGH, P. 2010. Enabling late-stage translation of regenerative medicine based products. © Pawanbir Singh.

TAN, T. E., PEH, G. S., FINKELSTEIN, E. A. & MEHTA, J. S. 2015. A practical model for economic evaluation of tissue-engineered therapies. Wiley Interdisciplinary Reviews: Systems Biology and Medicine, 7, 91-100.

VALLEJO-TORRES, L., STEUTEN, L., PARKINSON, B., GIRLING, A. J. & BUXTON, M. J. 2011. Integrating health economics into the product development cycle: a case study of absorbable pins for treating hallux valgus. Medical Decision Making, 31, 596-610.

VAN NIMWEGEN, K. J., LILFORD, R. J., VAN DER WILT, G. J. & GRUTTERS, J. P. 2017. HEADROOM BEYOND THE QUALITY-ADJUSTED LIFE-YEAR: THE CASE OF COMPLEX PEDIATRIC NEUROLOGY. International Journal of Technology Assessment in Health Care, 1-6.

Early methods (excluding HTA to inform developers)

ANDRONIS, L. 2015. Analytic approaches for research priority-setting: issues, challenges and the way forward. Expert review of pharmacoeconomics & outcomes research, 15, 745-754.

FASTERHOLDT, I., LEE, A., KIDHOLM, K., YDERSTRÆDE, K. B. & PEDERSEN, K. M. 2018. A qualitative exploration of early assessment of innovative medical technologies. BMC health services research, 18, 837.

HALL, P. S., MCCABE, C., BROWN, J. M. & CAMERON, D. A. 2010. Health economics in drug development: efficient research to inform healthcare funding decisions. European Journal of Cancer, 46, 2674-2680.

HARTZ, S. & JOHN, J. 2009. Public health policy decisions on medical innovations: What role can early economic evaluation play? Health Policy, 89, 184-192.

HUTTON, J., TRUEMAN, P. & HENSHALL, C. 2007. Coverage with evidence development: an examination of conceptual and policy issues. International journal of technology assessment in health care, 23, 425-432.

JÖNSSON, B. 2015. Bringing in health technology assessment and cost-effectiveness considerations at an early stage of drug development. Molecular oncology, 9, 1025-1033.

LEAL, J., WORDSWORTH, S., LEGOOD, R. & BLAIR, E. 2007. Eliciting expert opinion for economic models: an applied example. Value in Health, 10, 195-203.

MYERS, E., MCBROOM, A. J., SHEN, L., POSEY, R. E., GRAY, M. R. & SANDERS, G. D. 2012. Value-of-information analysis for patient-centered outcomes research prioritization. Report prepared by the Duke Evidence-based Practice Center. Patient-Centered Outcomes Research Institute.

MYERS, E., SANDERS, G. D., RAVI, D., MATCHAR, D., HAVRILESKY, L., SAMSA, G., POWERS, B., MCBROOM, A., MUSTY, M. & GRAY, R. 2011. Evaluating the

potential use of modeling and value-of-information analysis for future research prioritization within the evidence-based practice center program.

PALMER, S. & SMITH, P. C. 2000. Incorporating option values into the economic evaluation of health care technologies. Journal of Health Economics, 19, 755-766.

RETÈL, V. P., JOORE, M. A., LINN, S. C., RUTGERS, E. J. & VAN HARTEN, W. H. 2012. Scenario drafting to anticipate future developments in technology assessment. BMC research notes, 5, 442.

SCULPHER, M., DRUMMOND, M. & BUXTON, M. 1997. The iterative use of economic evaluation as part of the process of health technology assessment. Journal of Health Services Research, 2, 26-30.

VAN HARTEN, W. & RETÈL, V. 2016. Innovations that reach the patient: early health technology assessment and improving the chances of coverage and implementation. ecancermedicalscience, 10.

Early applied (excluding HTA to inform developers)

BAROCAS, D. A., BENSINK, M. E., BERRY, K., MUSA, Z., BODNAR, C., DANN, R. & RAMSEY, S. D. 2014. Economic evaluation of diagnostic localization following biochemical prostate cancer recurrence. International journal of technology assessment in health care, 30, 345-353.

BOSCH, J. L., BEINFELD, M. T., MULLER, J. E., BRADY, T. & GAZELLE, G. S. 2005. A Cost–Effectiveness Analysis of a Hypothetical Catheter-Based Strategy for the Detection and Treatment of Vulnerable Coronary Plaques with Drug-Eluting Stents. Journal of interventional cardiology, 18, 339-349.

BRIGGS, A., SCULPHER, M., BRITTON, A., MURRAY, D. & FITZPATRICK, R. 1998. The costs and benefits of primary total hip replacement: how likely are new prostheses to be cost-effective? International journal of technology assessment in health care, 14, 743-761.

BRODTKORB, T.-H. 2010. Cost-effectiveness analysis of health technologies when evidence is scarce. Linköping University Electronic Press.

DONG, H. & BUXTON, M. 2006. Early assessment of the likely cost-effectiveness of a new technology: a Markov model with probabilistic sensitivity analysis of computer-assisted total knee replacement. International Journal of Technology Assessment in Health Care, 22, 191-202.

DONG, H., COYLE, D. & BUXTON, M. 2007. Value of information analysis for a new technology: computer-assisted total knee replacement. International journal of technology assessment in health care, 23, 337-342.

DRANITSARIS, G. & LEUNG, P. 2004. Using decision modeling to determine pricing of new pharmaceuticals: the case of neurokinin-1 receptor antagonist antiemetics for cancer chemotherapy. International journal of technology assessment in health care, 20, 289-295.

FLOBBE, K., KESSELS, A. G., SEVERENS, J. L., BEETS, G. L., DE KONING, H. J., VON MEYENFELDT, M. F. & VAN ENGELSHOVEN, J. M. 2004. Costs and effects of

ultrasonography in the evaluation of palpable breast masses. International journal of technology assessment in health care, 20, 440-448.

GAULTNEY, J. G., REDEKOP, W. K., SONNEVELD, P., VAN DER HOLT, B. & UYL-DE GROOT, C. A. 2013. A modeling study to assess the potential therapeutic and economic value of risk-stratified treatment in multiple myeloma using molecular markers of adverse prognosis. Economic evaluations, 139.

GIRLING, A. J., FREEMAN, G., GORDON, J. P., POOLE-WILSON, P., SCOTT, D. A. & LILFORD, R. J. 2007. Modeling payback from research into the efficacy of left-ventricular assist devices as destination therapy. International Journal of Technology Assessment in Health Care, 23, 269-277.

HANDELS, R. L., JOORE, M. A., TRAN-DUY, A., WIMO, A., WOLFS, C. A., VERHEY, F. R. & SEVERENS, J. L. 2015. Early cost-utility analysis of general and cerebrospinal fluid-specific Alzheimer's disease biomarkers for hypothetical disease-modifying treatment decision in mild cognitive impairment. Alzheimer's & Dementia, 11, 896-905.

HESSEL, F. P. 2006. Economic evaluation of the artificial liver support system MARS in patients with acute-on-chronic liver failure. Cost Effectiveness and Resource Allocation, 4, 16.

HUANG, W., GAYDOS, C. A., BARNES, M. R., JETT-GOHEEN, M. & BLAKE, D. R. 2013. Comparative effectiveness of a rapid point-of-care test for detection of Chlamydia trachomatis among women in a clinical setting. Sex Transm Infect, 89, 108-114.

HUMMELINK, S., GERRITS, J., KOOL, L. S., ULRICH, D., ROVERS, M. & GRUTTERS, J. 2017. The merits of decision modelling in the earliest stages of the IDEAL framework: An innovative case in DIEP flap breast reconstructions. Journal of Plastic, Reconstructive & Aesthetic Surgery.

KOLOMINSKY-RABAS, P. L., KRIZA, C., DJANATLIEV, A., MEIER, F., UFFENORDE, S., RADELEFF, J., BAUMGÄRTEL, P., LEB, I., SEDLMAYR, M. & GAISER, S. 2016.

Health economic impact of a pulmonary artery pressure sensor for heart failure telemonitoring: a dynamic simulation. Telemedicine and e-Health, 22, 798-808.

LAMBOOIJ, M. S. & HUMMEL, M. J. 2013. Differentiating innovation priorities among stakeholder in hospital care. BMC medical informatics and decision making, 13, 91.

LINDENBERG, M. A., RETÈL, V. P., VAN DEN BERG, J. H., FOPPEN, M. H. G., HAANEN, J. B. & VAN HARTEN, W. H. 2018. Treatment With Tumor-infiltrating Lymphocytes in Advanced Melanoma: Evaluation of Early Clinical Implementation of an Advanced Therapy Medicinal Product. Journal of immunotherapy (Hagerstown, Md.: 1997), 41, 413.

MIQUEL-CASES, A., STEUTEN, L. M., RETÈL, V. P. & VAN HARTEN, W. H. 2015. Early stage cost-effectiveness analysis of a BRCA1-like test to detect triple negative breast cancers responsive to high dose alkylating chemotherapy. The Breast, 24, 397-405.

MIQUEL-CASES, A., RETÈL, V. P., LEDERER, B., VON MINCKWITZ, G., STEUTEN, L. M. & VAN HARTEN, W. H. 2016a. Exploratory Cost-Effectiveness Analysis of Response-Guided Neoadjuvant Chemotherapy for Hormone Positive Breast Cancer Patients. PloS one, 11, e0154386.

MIQUEL-CASES, A., RETÈL, V. P., VAN HARTEN, W. H. & STEUTEN, L. M. 2016b. Decisions on further research for predictive biomarkers of high-dose alkylating chemotherapy in triple-negative breast cancer: a value of information analysis. Value in health, 19, 419-430.

MORRIS, A. M., FLOWERS, C. R., MORRIS, K. T., SCHMIDT, W. A., POMMIER, R. F. & VETTO, J. T. 2003. Comparing the Cost-Effectiveness of the Triple Test Score to Traditional Methods for Evaluating Palpable Breast Masses. Medical care, 41, 962-971.

ØSTERGAARD, S. & MØLDRUP, C. 2010. Anticipated outcomes from introduction of 5-HTTLPR genotyping for depressed patients: An expert Delphi analysis. Public health genomics, 13, 406-414.

PIRC, E., PECCHIA, L., REBERŠEK, M., SERŠA, G., SNOJ, M., GROŠELJ, A. & MIKLVAVČIČ, D. 2018. Study design of a medical device pre-market assessment: a case study on electrochemotherapy. Slovenian Medical Journal, 87.

RETEL, V. P., JOORE, M., DRUKKER, C., BUENO-DE-MESQUITA, J., KNAUER, M., VAN TINTEREN, H., LINN, S. & VAN HARTEN, W. H. 2013. Prospective cost-effectiveness analysis of genomic profiling in breast cancer. European journal of cancer, 49, 3773-3779.

RETÈL, V. P., STEUTEN, L. M., FOPPEN, M. H. G., MEWES, J. C., LINDENBERG, M. A., HAANEN, J. B. & VAN HARTEN, W. H. 2018. Early cost-effectiveness of tumor infiltrating lymphocytes (TIL) for second line treatment in advanced melanoma: a model-based economic evaluation. BMC cancer, 18, 895.

ROOS, J. B., DOSHI, S. N., KONORZA, T., PALACIOS, I., SCHREIBER, T., BORISENKO, O. V. & HENRIQUES, J. P. 2013. The cost-effectiveness of a new percutaneous ventricular assist device for high-risk PCI patients: mid-stage evaluation from the European perspective. Journal of medical economics, 16, 381-390.

SCHWANDER, B. 2014. Early health economic evaluation of the future potential of next generation artificial vision systems for treating blindness in Germany. Health economics review, 4, 27.

STEUTEN, L., VALLEJO-TORRES, L., YOUNG, T. & BUXTON, M. 2008. Transferability of economic evaluations of medical technologies: a new technology for orthopedic surgery. Expert review of medical devices, 5, 329-336.

STEUTEN, L., VALLEJO-TORRES, L., BASTIDE, P. & BUXTON, M. 2009. Analysing uncertainty around costs of innovative medical technologies: the case of fibrin sealant (QUIXIL®) for total knee replacement. Health policy, 89, 46-57.

SVENSSON, J., GHATNEKAR, O., LINDGREN, A., LINDVALL, O., NORRVING, B., PERSSON, U. & KOKAIA, Z. 2012. Societal value of stem cell therapy in stroke–a modeling study. Cerebrovascular Diseases, 33, 532-539.

TAPPENDEN, P., SACCARDI, R., CONFAVREUX, C., SHARRACK, B., MURARO, P., MANCARDI, G., KOZAK, T., FARGE-BANCEL, D., MADAN, J. & RAFIA, R. 2010. Autologous haematopoietic stem cell transplantation for secondary progressive multiple sclerosis: an exploratory cost-effectiveness analysis. Bone marrow transplantation, 45, 1014.

TERJESEN, C. L., KOVALEVA, J. & EHLERS, L. 2017. Early Assessment of the Likely Cost Effectiveness of Single-Use Flexible Video Bronchoscopes. PharmacoEconomics-Open, 1, 133-141.

VAN TIL, J. A., RENZENBRINK, G. J., GROOTHUIS, K. & IJZERMAN, M. J. 2006. A preliminary economic evaluation of percutaneous neuromuscular electrical stimulation in the treatment of hemiplegic shoulder pain. Disability and rehabilitation, 28, 645-651.

WEBER, S. A., PIETZSCH, M., BESTARD, O., GRINYO, J. M., VIKLICKY, O. & REINKE, P. 2016. Investigating the health-economic profiles of biomarker-driven immunosuppresion (BIO-DrIM) following solid organ transplantation. Advances in Precision Medicine, 1, 48-65.

WINDT, T. S., SOREL, J. C., VONK, L. A., KIP, M., IJZERMAN, M. J. & SARIS, D. B. 2016. Early health economic modelling of single-stage cartilage repair. Guiding implementation of technologies in regenerative medicine. Journal of tissue engineering and regenerative medicine.

WONG, W. B., RAMSEY, S. D., BARLOW, W. E., GARRISON, L. P. & VEENSTRA, D. L. 2012. The value of comparative effectiveness research: projected return on investment of the RxPONDER trial (SWOG S1007). Contemporary clinical trials, 33, 1117-1123.

Combined early methods and applied papers (excluding HTA to inform developers)

ANDRONIS, L. 2013. Analytic approaches for informing research funding decisions: an exploration of their role and value using case studies of cancer clinical trials. University of Birmingham.

BALLINI, L., MINOZZI, S., NEGRO, A., PIRINI, G. & GRILLI, R. 2010. A method for addressing research gaps in HTA, developed whilst evaluating robotic-assisted surgery: a proposal. Health research policy and systems, 8, 27.

BENNETTE, C. 2015. Prioritizing research: the use of risk prediction, value of information analyses, and portfolio evaluation to improve public investments in cancer clinical trials.

BENNETTE, C. S., VEENSTRA, D. L., BASU, A., BAKER, L. H., RAMSEY, S. D. & CARLSON, J. J. 2016. Development and evaluation of an approach to using value of information analyses for real-time prioritization decisions within SWOG, a large cancer clinical trials cooperative group. Medical Decision Making, 36, 641-651.

BOYD, K. A. 2012. Employing early decision analytic modelling to inform economic evaluation in health care: theory & practice. University of Glasgow.

CARLSON, J. J., THARIANI, R., ROTH, J., GRALOW, J., HENRY, N. L., ESMAIL, L., DEVERKA, P., RAMSEY, S. D., BAKER, L. & VEENSTRA, D. L. 2013. Value-of-information analysis within a stakeholder-driven research prioritization process in a US setting: an application in cancer genomics. Medical Decision Making, 33, 463-471.

DEGELING, K., SCHIVO, S., MEHRA, N., KOFFIJBERG, H., LANGERAK, R., DE BONO, J. S. & IJZERMAN, M. J. 2017. Comparison of timed automata with discrete event simulation for modeling of biomarker-based treatment decisions: an illustration for metastatic castration-resistant prostate cancer. Value in health, 20, 1411-1419.

DOUMA, K. F., KARSENBERG, K., HUMMEL, M. J., BUENO-DE-MESQUITA, J. M. & VAN HARTEN, W. H. 2007. Methodology of constructive technology assessment in health care. International Journal of Technology Assessment in Health Care, 23, 162-168.

HALL, P. S., MITCHELL, E. D., SMITH, A. F., CAIRNS, D. A., MESSENGER, M., HUTCHINSON, M., WRIGHT, J., VINALL-COLLIER, K., CORPS, C. & HAMILTON, P. 2018. The future for diagnostic tests of acute kidney injury in critical care: evidence synthesis, care pathway analysis and research prioritisation.

MILES, K. A. 2001. An approach to demonstrating cost-effectiveness of diagnostic imaging modalities in Australia illustrated by positron emission tomography. Journal of Medical Imaging and Radiation Oncology, 45, 9-18.

NOTA, I. 2009. Technology assessment of innovative operating room technology. University of Twente.

RETEL, V., RETÈL, V. P., HUMMEL, J. M. & VAN HARTEN, W. H. 2008. Early phase Technology Assessment of nanotechnology in oncology. Tumori, 94, 284-290.

RETÈL, V. P., BUENO-DE-MESQUITA, J. M., HUMMEL, M. J., VAN DE VIJVER, M. J., DOUMA, K. F., KARSENBERG, K., VAN DAM, F. S., VAN KRIMPEN, C., BELLOT, F. E. & ROUMEN, R. M. 2009. Constructive Technology Assessment (CTA) as a tool in coverage with evidence development: the case of the 70-gene prognosis signature for breast cancer diagnostics. International journal of technology assessment in health care, 25, 73-83.

RETÈL, V. P., GRUTTERS, J. P., VAN HARTEN, W. H. & JOORE, M. A. 2013. Value of research and value of development in early assessments of new medical technologies. Value in health, 16, 720-728.

YAO, G. L., NOVIELLI, N., MANASEKI-HOLLAND, S., CHEN, Y.-F., VAN DER KLINK, M., BARACH, P., CHILTON, P. J. & LILFORD, R. J. 2012. Evaluation of a predevelopment service delivery intervention: an application to improve clinical handovers. BMJ Qual Saf, bmjqs-2012-001210.

Appendix 3 – Studies with explicit frameworks characterising development-focused HTA

First author	Target audience	Timing	Specific decision HTA to inform	Available evidence
Pietzsch (2008) Early technology assessment of new medical devices	"Classical HTA" informs "regulators, payers and patients"		Informs "market clearance, payment and usage of a technology" Authors include these decisions under "aims" of HTA "Assess safety, effectiveness, and cost-effectiveness" "Limited on no influence on clinical performance'	"Usually evidence from clinical studies performed with the new technology"
	"Early HTA" informs "manufacturers and investors"		Informs about "design and management of a technology, as well as regulatory and reimbursement strategy" Authors include these decisions under "aims" of HTA "Assess *likely* safety, effectiveness and cost-effectiveness" "Potentially significant influence on (future) clinical performance"	"Evidence from early bench and animal testing, early clinical experience and previous generations of the technology"
Ijzerman (2011) Early assessment of medical technologies to inform product development and market access: A review of methods and applications	"Mainstream HTA" informs "Government, industry and clinical and basic research centres"	"At market access and pricing stage":	Government – "Do we approve new products for market access and for what price?" Industry – "Develop a value dossier. What is the best market access strategy? What is best possible price given regulatory constraints?" Clinical and basic research centres – "How do we get timely access to the new products and how do we manage and monitor quality assurance?"	
		At clinical research stage:	Government – "Should we use public research funding for clinical research on new products?" Industry – "Where and how to launch the product. How to organise the clinical research strategy. Develop the economic evidence alongside." Clinical and basic research centres – "What is the clinical evidence for new products? How to design and manage our clinical research?"	

First author	Target audience	Timing	Specific decision HTA to inform	Available evidence
	"Early HTA" informs "Government, industry and clinical and basic research centres"	At translational research stage:	Government – "Should we invest in product development for economic growth and health benefits? Is there a market failure?" Industry – "Should we invest in R&D and target our products to specific patient groups?"" Clinical and basic research centres – "Should we apply for public funding and in what consortia? What is our research focus and does the project fit our portfolio?"	
	"Very early HTA" informs "Government, industry and clinical and basic research centres"	At basic research stage:	Government – "Should we use public resources for further research? To which consortia should it be awarded?" Industry – "Should we invest in future research, expand our research portfolio or partner with other public/private partners? How to obtain public funding? Clinical and basic research centres – "Should we apply for public funding and in what consortia? What is our research focus and does the project fit our portfolio?"	

HTA – Health Technology Assessment, R&D – Research and development " " indicates that the quote is directly from the source article

Appendix 4 – Informative papers distinguishing development and reimbursement-focused HTA

First author (Date) Title	Target audience	Underlying objectives	Specific decision HTA to inform	Core decision rule	Available evidence	Timing	Clinical decision space	Business model	Resources for analysis
Annemans – (2000) Early modelling for assessing health and economic outcomes of drug therapy	"Pharmaceutical industry" "Healthcare decision-makers and payers" NB Annemans makes the point that the perspective of the model should be the payers regardless of who the model is seeking to inform		Early modelling informs pharmaceutical industry – 'Go/No go decisions", "priority-setting decisions", "choice of indication", positioning, comparators, length of follow-up and other elements in the further development of drugs", importance of "a given parameter" "influence future price and reimbursement status" Late HTA informs reimbursement decision-makers (payers) "about allocating scarce resources", "decisions about formularies" and "optimal medical treatment guidelines", "revision of price and reimbursement decisions", "budgetary implications"		Early models have "very scarce data" "Anticipated clinical profile" based on "management desires" or "pharmacodynamic properties" Evidence for early models – clinical data for comparators "from literature" with "very careful analysis required", resource use from literature, databases or expert opinion – can compromise here		Recognises decision space over jurisdiction, indication, comparator, position in pathway and dosage Time horizon of the model may also be varied Deal with using multiple scenario analysis		Early models have "limited budgets and limited timelines"

First author (Date) Title	Target audience	Underlying objectives	Specific decision HTA to inform	Core decision rule	Available evidence	Timing	Clinical decision space	Business model	Resources for analysis
Cosh (2007) Investing in new medical technologies: a decision framework	"A company considering the development of a new medical technology"		Whether to invest in a new technology or continue to invest. "To avoid investing in those devices, which can never be cost-effective"	Continue to invest if headroom and return on investment are positive	Since headroom is a "rough and ready method to decide whether to continue with development, less formal and inexpensive methods (of expert elicitation) will often be fit for purpose"	"Before substantial investments are made" "As research progresses, estimates of costs and effectiveness can be updated"	Suggests scenario analysis with different levels of product performance as last stage in framework	Third party payer with known threshold assumed in headroom analysis. Consumers would each have "a different willingness to pay".	
Hartz (2008) Contribution of economic evaluation to decision making in early phases of product development: A methodological and empirical review	"Industry"	Maximise return on investment	"Pre-clinical preliminary market assessments" "Go/no-go decisions, identification of potentially successful projects" "R&D portfolio management" "First estimations of pricing and reimbursement scenarios" "Development of future trial design" "Identify gaps in evidence needed"		"Literature reviews, claims data or national health surveys" for pre-clinical market assessments. "First data from Phase I/II trials for Go/No-go decisions"	"Early phases of product development"	"Evaluation of the cost-effectiveness at different pricing scenarios, patient populations and indications can be carried out".		

First author (Date) Title	Target audience	Underlying objectives	Specific decision HTA to inform	Core decision rule	Available evidence	Timing	Clinical decision space	Business model	Resources for analysis
Vallejo-Torres (2008) Integrating health economic modelling in the product development cycle of medical devices: A Bayesian approach	"Medical device companies"		"To avoid investing in a technology that could never be cost-effective" "To prioritise between several competing possibly cost-effective concepts or prototypes" To identify "from early stages of development those parameters that have the largest impact on the likely cost-effectiveness of the product" "To help companies reduce their failure rates"	Uses threshold value in given jurisdiction to determine whether technology likely to be cost effective but acknowledges that these models should only inform a wider commercial perspective	"Early stage" – "Based on available evidence concerning the current technology that the new device aims to substitute" and "expert opinion and/or assumptions" "Mid-stage" – "Observational studies" "initial cost estimates" "prior beliefs .. elicited from a group of experts" "Final-stage" – all available evidence	Three stage process –early-stage, mid-stage and late-stage	"Different types of interventions, different clinical settings or different clinical indications"		Early in the development process time to perform analyses may be scarce Use less formal/time-consuming methods of expert elicitation early in development process

First author (Date) Title	Target audience	Underlying objectives	Specific decision HTA to inform	Core decision rule	Available evidence	Timing	Clinical decision space	Business model	Resources for analysis
Mikudina (2013) Early medical technology assessments (MTA) of medical devices and tests	Classical Medical Technology Assessment (MTA) informs "regulators and payers" "Early MTA" informs "Researchers and manufacturers"		Classical MTA To support policymakers to "assess the overall value of a drug, medical device or diagnostic test". Early MTA "To support researchers and manufacturers "to make better decisions about further development, the regulatory and reimbursement strategy and allocating public support for new technologies" "To improve the device during the development process to produce the most beneficial medical technology for society" "To help design future trials"		Classical MTA "after large clinical trials, when clinical and cost-effectiveness data are available" Early MTA in early and mid-stages as identified by Vallejo-Torres. Evidence in early stage "assumptions", in mid-stage "some evidence from pre-clinical studies"	Classical MTA "at the end of the development process" Early MTA "in the early phases" of the development process			Headroom method is "a quick and easy model"

First author (Date) Title	Target audience	Underlying objectives	Specific decision HTA to inform	Core decision rule	Available evidence	Timing	Clinical decision space	Business model	Resources for analysis
Rogowski (2016) Translational health economics	"Translational Health Economics" bridges gap between decisions to "invest in development of a technology" and to "fund and use" it in clinical practice.		To "analyse "value" for a potential health technology" and decide whether "to initiate a translation process" To 'analyse "uncertainty" and to decide which methods to use to generate evidence about value "early on" To "analyse barriers of information and motivation in the cooperative process of translation as well as institutions to overcome these"			"Concept stage" "Development stage" "Translation stage"	Scenario analysis suggested given different applications and levels of uptake. "Wide methods of value capture suggested for different stakeholder perceptions of value		
Lal (2014) The overarching framework of translation and integration into health care: a case for the LAL model	"Academic-industrial complex" "Decision makers"					Three stages: T1 -basic to applied research T2 – translation to "industrial application" T3 – "implement in healthcare systems through best practice guidelines and health policy"			

First author (Date) Title	Target audience	Underlying objectives	Specific decision HTA to inform	Core decision rule	Available evidence	Timing	Clinical decision space	Business model	Resources for analysis
Markiewicz (2014) Medical Devices Early Assessment Methods: Systematic Literature Review	"Manufacturers" "Policy makers" "Decision makers on coverage and reimbursement" "Varied"		"Potential applications or improvement directions" "End -user perspectives in further development"		Increases throughout development process Need to acknowledge and estimate impact of uncertainty	Iterative due to technology being a moving target			
Girling (2015) Headroom approach to device development: current and future directions	Developers	Maximising "a product's ultimate profitability"	"Reality check on the viability of the device" "Support product development decisions using a real options approach" "To contribute to a pricing policy" "Termination of development"	Optimistic assessment of potential clinical value at outset.	Little or no evidence available at outset of development. Uncertainty can be modelled using simple probability distributions.	Headroom approach from early concept to pre-market stage, iterative and increasing in complexity Less uncertainty later in development.	Could option to stop development be seen as an aspect of decision space?		"Simple approach to make rapid decisions at the start of product development" Headroom is a "simple shared tool to parties in commercial negotiation"
Buisman (2016b) – The Early Bird Catches the Worm: Early cost-effectiveness analysis of new medical tests	"Medical test developer"	"Return on Investment"	"To decide about further development of medical tests" "To set realistic performance-price goals" To "design and manage reimbursement strategies"	Invest if return on investment is positive	"Much less data available for early CEAs"	"Throughout the development of a test, new data and ideas may emerge" Early CEAs are "much more iterative"	Position in pathway, indication and population may not be settled. Use of scenario analysis to look at alternatives		"It might seem resource intensive at first"

First author (Date) Title	Target audience	Underlying objectives	Specific decision HTA to inform	Core decision rule	Available evidence	Timing	Clinical decision space	Business model	Resources for analysis
Ijzerman (2017) Emerging Use of Early Health Technology Assessment in Medical Product Development: A scoping review of the literature	"Industry"	"Revenue maximisation"	"Strategic R&D decisions" "Preclinical market assessment" "Portfolio decisions" "Clinical trial design" "Market access and pricing strategies"						

CEA – Cost effectiveness analysis. R&D - Research and development. "quote marks" indicate that the contents are a direct quote from the artic

Appendix 5 – Features of applied and combined methods and applied papers in DF-HTA

First author	Target audience	Underlying objective	Stance of analysis	Decisions HTA designed to inform	Core decision rule	Evidence specific to technology available?	Burden of proof	Timing	Clinical decision space	Business model	Resources for analysis
Brandes (2015)	Manufacturers	Not stated	Unclear	Reimbursement / regulatory strategy Technology design	Value-based pricing uses German efficiency frontier	None	Not discussed	Concept	Single jurisdiction Multiple indications, sub-groups	Reimbursement	Constrained: German Federal Ministry of Education and Research
Braz (2013)	Developers	Not stated	Unclear	Reimbursement / regulatory strategy	Not stated	Some	Not discussed	Phase I trial ongoing	Two jurisdictions Single indication Multiple positions in pathway	Reimbursement	Masters Thesis
Breteler (2012)	Investors	Not stated	Unclear	Reimbursement / regulatory strategy Technology design	Not stated	None	Not discussed	Prototype	Single jurisdiction Single indication Multiple positions in pathway	Reimbursement	Masters Thesis
Buisman (2016a)	Developers	Not stated	Unclear	Reimbursement / regulatory strategy Go/no-go - ongoing investment	Not stated	None	Not discussed	Prototype	Single jurisdiction Single indication Multiple positions in pathway	Reimbursement and other business models	Center for Translational Molecular Medicine and project specific grants

First author	Target audience	Underlying objective	Stance of analysis	Decisions HTA designed to inform	Core decision rule	Evidence specific to technology available?	Burden of proof	Timing	Clinical decision space	Business model	Resources for analysis
Buisman (2016b)	Unclear	Not stated	Unclear	Reimbursement / regulatory strategy Technology design Go/no-go - ongoing investment	Threshold of €20,000 used	Some	Not discussed	Prototype	Multiple tests Single jurisdiction Single indication Single position in pathway	Reimbursement	TRACER project (Center for Translational Molecular Medicine)
Cao (2013)	Developers	Maximise commercial return	Positive	Reimbursement / regulatory strategy Go/no-go - ongoing investment	Proceed if sufficient return on investment is likely	None	Not discussed	Concept	Single jurisdiction Single indication Multiple positions in pathway	Reimbursement and other business models	Center for Translational Molecular Medicine and project specific grants
Carr (2013)	Developer and policy-maker	Maximise societal and commercial return	Unclear	Evidence generation strategy Go/no-go - ongoing investment	Not stated	None	Not discussed	Pre Phase I trials	Single jurisdiction Single indication Multiple comparators, population sub-groups	Reimbursement	PhD thesis
Chapman (2012)	Medical device developers	Not stated	Positive	Reimbursement / regulatory strategy Go/no-go - ongoing investment	Net present value greater than 0	None	Not discussed	Concept	Multiple jurisdiction Multiple indication Multiple positions in pathway	Reimbursement and other business models	Constrained – headroom in a day PhD thesis
Craven (2011)	Developers	Not stated	Unclear	Reimbursement / regulatory strategy	Threshold of £20,000-£30,000 used	None	Not discussed	Prototype	Single jurisdiction Single indication Single position in pathway	Reimbursement	Constrained – 2 week decision model MATCH UK and voucher scheme

First author	Target audience	Underlying objective	Stance of analysis	Decisions HTA designed to inform	Core decision rule	Evidence specific to technology available?	Burden of proof	Timing	Clinical decision space	Business model	Resources for analysis
Davey (2011)	Health service and industry decision makers	Not stated	Unclear	Reimbursement / regulatory strategy	Not stated	None	Not discussed	Prototype and concept	Single jurisdiction Multiple indication Multiple positions in pathway	Reimbursement and other business models	MATCH UK
de Graaf (2015)	Research team	Maximise health (minimise burden)	Unclear	Portfolio management	Maximum score on six criteria decision analysis	None	Not discussed	Concept	Multiple technologies Single jurisdiction Single indication Single position in pathway	Reimbursement and other business models	Center for Translational Molecular Medicine and project specific grants
de Graaf (2018)	Research team	Maximise health (minimise burden)	Unclear	Reimbursement / regulatory strategy Portfolio management	Range of willingness to accept thresholds presented	None	Not discussed	Analytic validity has been established	Single jurisdiction Single indication Single position in pathway	Reimbursement and other business models	Constrained – downstream modelling only Center for Translational Molecular Medicine and project specific grants
Gantner-Bar (2011)	Healthcare players	Not stated	Unclear	Technology design Go/no go – continuing investment	Not applicable	Not applicable	Not discussed	Concept	Single jurisdiction Single indication Single position in pathway	Reimbursement under the German insurance system	German Federal Ministry of Education and Research

First author	Target audience	Underlying objective	Stance of analysis	Decisions HTA designed to inform	Core decision rule	Evidence specific to technology available?	Burden of proof	Timing	Clinical decision space	Business model	Resources for analysis
Gaultney (2011)	Developers and policy-makers	Maximise societal and commercial return	Unclear	Reimbursement / regulatory strategy Evidence generation strategy	Threshold analysis based on cost neutrality	Some	Not discussed	Prototype	Single jurisdiction Single indication Single position in pathway	Reimbursement in the specific jurisdiction	Grant funding from PamGene
Groothius-Oudshoorn (2014)	Developers? Policy-makers as developers given it is a screening programme?	Not stated	Unclear	Technology design	Not stated	None	Not discussed	Preclinical development phase	Single jurisdiction Single indication Single position in pathway	Reimbursement - population screening programme	Not stated
Haakma (2014)	Developers	Not stated	Unclear	Evidence generation strategy Technology design	Not stated	Some	Not discussed	Prototype	Single jurisdiction Single indication Single position in pathway	Not applicable	Not stated
Hilgerink (2011)	Developer	Maximise commercial return	Unclear	Reimbursement / regulatory strategy Technology design	Not stated	Some	Not discussed	Prototype	Single jurisdiction Single indication Multiple positions in pathway	Not applicable	Project specific funding
Hjelmgren (2006)	Unclear	Not stated	Unclear	Evidence generation strategy Technology design	Not stated	Some	Not discussed	Technology has been tried in small clinical trial	Single jurisdiction Single indication Single position in pathway	Reimbursement	Project specific funding

First author	Target audience	Underlying objective	Stance of analysis	Decisions HTA designed to inform	Core decision rule	Evidence specific to technology available?	Burden of proof	Timing	Clinical decision space	Business model	Resources for analysis
Hummel (2012)	Biomedical engineers	Maximise health	Positive (aim to support future develop ment)	Evidence generation strategy Go/no-go - ongoing investment	Continue to invest if likely to be cost-effective	Some	Not discussed	Non-fusion surgery is a new treatment	Single jurisdiction Single indication Single position in pathway	Reimbursement	Not stated
Huygens (2016)	Developers, clinicians and payers	Maximise societal and commercial return	Unclear	Technology design	Not stated	None	Not discussed	Concept	Single jurisdiction Single indication Single position in pathway	Reimbursement	Netherlands Cardio Vascular Research Initiative (CVON) Ivalve consortium
Kenter (2015)	Developers, payers and users	Not stated	Unclear	Reimbursement / regulatory strategy	Not stated	None	Not discussed	Prototype	Single jurisdiction Single indication Multiple positions in pathway	Reimbursement	Masters Thesis
Khoudigian – Sinani (2017)	Developer, investor, manufacturer, healthcare system and individual patient	Not stated	Unclear	Reimbursement / regulatory strategy	Continue to invest if likely to be cost-effective	Some	Not discussed	Pre-market	Single jurisdiction Single indication Multiple populations	Private payer and patient perspectives	The MITACS Accelerate Program (Federal Government and Proteocyte Diagnostic Inc)

First author	Target audience	Underlying objective	Stance of analysis	Decisions HTA designed to inform	Core decision rule	Evidence specific to technology available?	Burden of proof	Timing	Clinical decision space	Business model	Resources for analysis
Kip (2016)	Developers	Not stated	Unclear	Reimbursement / regulatory strategy Technology design	Continue development if incremental cost effectiveness ratio lower than acceptable threshold	Some	Not discussed	Test not currently in clinical practice	Single jurisdiction Single indication Single position in pathway Multiple levels of test performance and strategies	Reimbursement	PhD thesis
Kluytmans (2019)	Developers	Not stated	Positive	Go/no-go - ongoing investment	Not stated	Some	Not discussed	Prototype	Multiple jurisdictions Multiple surgical applications	Reimbursement	ZonMw Innovative Medical Devices Initiative Program
Knuttell (2017)	Developers	Not stated	Unclear	Technology design	Not stated	Some	Not discussed	Prototype	Multiple jurisdictions Single indication Single position in pathway	Reimbursement	Center for Translational Molecular Medicine
Koerber (2013)	Developers	Not stated	Unclear	Reimbursement / regulatory strategy Evidence generation strategy Go/no-go - ongoing investment	Not stated	Some	Not discussed	Unclear	Single jurisdiction Single indication Single position in pathway	Reimbursement or direct contract with hospital	German Federal Ministry of Education and Research

First author	Target audience	Underlying objective	Stance of analysis	Decisions HTA designed to inform	Core decision rule	Evidence specific to technology available?	Burden of proof	Timing	Clinical decision space	Business model	Resources for analysis
Kolominsky-Rabas (2015)	Manufacturers and healthcare decision makers	Not stated	Unclear	Go/no-go - ongoing investment Technology design	Not stated	Some	Not discussed	Prototype	Single jurisdiction Single indication Single position in pathway	Reimbursement under the German insurance system	National Cluster of Excellence "Medical Technology - Medical Valley EMN"
Koning (2012)	Developer	Maximise commercial return	Positive	Reimbursement / regulatory strategy Technology design	Not stated	Some	Not discussed	Prototype	Multiple jurisdictions Multiple indications Multiple settings	Fluid	Masters Thesis
Latimer (2011)	Developers	Maximise health	Positive	Go/no-go - ongoing investment	Continue development if incremental cost effectiveness ratio lower than acceptable threshold	None	Not discussed	Concept	Single jurisdiction Single indication Single position in pathway	Reimbursement	Constrained - UK Department of Health
Liu (2013)	Developer	Not stated	Positive	Technology design Go/no-go - ongoing investment	Not stated	None	Not discussed	Concept	Multiple intervention Single jurisdiction Single indication Multiple population sub-groups	Reimbursement model tested	Phillips research

First author	Target audience	Underlying objective	Stance of analysis	Decisions HTA designed to inform	Core decision rule	Evidence specific to technology available?	Burden of proof	Timing	Clinical decision space	Business model	Resources for analysis
Luime (2016)	Developers	Not stated	Not stated	Reimbursement / regulatory strategy Technology design	Adopt service change if incremental cost effectiveness ratio lower than acceptable threshold	Some	Not discussed	Tests under development as part of TRACER project (CTMM)	Single jurisdiction Single indication Single position in pathway Multiple tests, performance levels and population sub-groups	Reimbursement	Center for Translational Molecular Medicine and project specific grants
Markiewicz (2017b)	Developers	Maximise return on investment	Positive	Reimbursement / regulatory strategy Technology design	Not stated	Not applicable	Not discussed	Targeting product and early proof of concept stages	Single jurisdiction Single indication Multiple settings, populations and positions in pathway	Fluid	Center for Translational Molecular Medicine
Markiewicz (2016)	Developers	Maximise return on investment	Positive	Portfolio management	Maximise headroom and financial return on investment	None	Not discussed	Concept Prototype	Single jurisdiction Single indication Single position in pathway	Reimbursement model tested	Constrained - Center for Translational Molecular Medicine
McAteer (2007)	Developers	Maximise return on investment	Positive	Go/no-go - ongoing investment	Continue to invest if headroom sufficiently high	None	Not discussed	Concept	Single jurisdiction Multiple indications Single position in pathway	Reimbursement model tested	Constrained - MATCH UK and project specific funding

First author	Target audience	Underlying objective	Stance of analysis	Decisions HTA designed to inform	Core decision rule	Evidence specific to technology available?	Burden of proof	Timing	Clinical decision space	Business model	Resources for analysis
Pietzsch (2008)	Investors and developers	Maximise financial and societal return on investment	Positive	Go/no-go - ongoing investment Technology design	Continue projects which best meet thresholds for risk set by the firm	None	Not discussed	Preprototype	Single jurisdiction Single indication Single position in pathway	Not specified. Model assumes value of innovation is a function of its impact upon outcomes	Project specific funding
Postmus (2012)	Investors and analysts	Maximise return on investment	Positive	Technology design	Prioritise projects with highest headroom and financial return on investment	Some	Not discussed	Analytic validity has been established	Single jurisdiction Single indication Multiple positions in pathway	Reimbursement model tested	Center for Translational Molecular Medicine and project specific grants
Rejon-Parilla (2014)	Developer (from discussion)	Maximise societal and commercial return	Not stated	Technology design	Not stated	Some	Not discussed	Biomarker has been associated with differential response but not used to guide dosing	Single jurisdiction Single indication Single position in pathway	Reimbursement	None
Singh (2010)	Industry	Maximise return on investment	Positive	Go/no-go - ongoing investment	Continue project if return on investment positive	None	Not discussed	Unclear	Single jurisdiction Single indication Single position in pathway	Reimbursement used as example of way to market	PhD thesis

First author	Target audience	Underlying objective	Stance of analysis	Decisions HTA designed to inform	Core decision rule	Evidence specific to technology available?	Burden of proof	Timing	Clinical decision space	Business model	Resources for analysis
Tan (2015)	Developers	Maximise return on investment	Positive	Go/no-go - ongoing investment	Pursue if economically viable	Some	Not discussed	Pre Phase I trials	Single jurisdiction Single indication Single position in pathway	Reimbursement model tested	Constrained – cost-minimisation
Vallejo-Torres (2011)	Developers	Maximise return on investment	Positive	Evidence generation strategy Go/no-go - ongoing investment	Continue project if return on investment positive	Early stage-none. Mid-stage - some	Discussion considers usefulness	Concept Prototype	Single jurisdiction Single indication Single position in pathway	Reimbursement used as example of way to market	Constrained – headroom MATCH UK
van de Wetering (2012)	Developers	Maximise commercial return	Positive	Reimbursement / regulatory strategy Evidence generation strategy Technology design Go/no-go - ongoing investment	Development proceed if cost below headroom	None	Not discussed	Prototype	Multiple jurisdictions Multiple indications Multiple positions in pathway, settings and populations	Reimbursement model tested	PhD thesis
van Nimwegen (2017)	Unclear	Not stated	Unclear	Go/no-go - ongoing investment	Potential for development if cost below headroom	None	Not discussed	Concept	Single jurisdiction Multiple indications Single position in pathway	Willingness to pay elicited - private payer market	Not stated

First author	Target audience	Underlying objective	Stance of analysis	Decisions HTA designed to inform	Core decision rule	Evidence specific to technology available?	Burden of proof	Timing	Clinical decision space	Business model	Resources for analysis
Vilsbøll (2018)	Unclear	Not stated	Unclear	Evidence generation strategy Technology design Go/no-go - ongoing investment	Not stated	Some	Not discussed	Clinical trials of the technology on going	Single jurisdiction Single indication Single position in pathway	Reimbursement	No funding received
Weiser (2013)	Unclear	Not stated	Unclear	Reimbursement / regulatory strategy Go/no-go - ongoing investment	Not stated	None	Not discussed	Prototype	Multiple jurisdictions Specific indication Multiple positions in the pathway and populations	Not applicable	Masters Thesis
Wissing (2012)	Developers	Maximise commercial return	Positive	Technology design	Not applicable	None	Not discussed	Technology is in development	Single jurisdiction Single indication Single position in pathway	Not clear	Masters Thesis

Appendix 6 – Categorisation process for methods of DF-HTA

Methods from review articles	Methods useful in DF-HTA included in Table 5-2	Notes
Methods from Markiewicz et al (2014)		
Qualitative		
Literature review/analysis	**Literature review**	Included
Peer review		Excluded as relevance not clear
User profiles building	**Qualitative methods of user interaction**	Included as Qualitative methods of user interaction
Focus groups		Included as Qualitative methods of user interaction
Interviews (e.g. experts)	**Expert opinion and expert elicitation**	Included as Expert opinion and expert elicitation
Informal discussions		Included as Expert opinion and expert elicitation
Qualitative weighing of relevant factors		Excluded as unclear what method involves
Use cases writing		Included as Qualitative methods of user interaction
Key informant interviews		Included as Qualitative methods of user interaction
Strategic planning methods: PEST, SWOT		Excluded – as strategic planning is primarily a commercial activity although these methods may be useful to structure qualitative methods of user interaction. Discussed under qualitative methods of user interaction.
Soft-systems methodology		Excluded - research and development process
Expert panels/elicitation		Included as Expert opinion and expert elicitation
Technology profiling/uncertainty profile and evidence profile)		Excluded - research and development process
Workshops		Included as Qualitative methods of user interaction
Surveys		Included as Qualitative methods of user interaction
Research and development portfolio management		Excluded - other commercial activities
Brainstorming sessions		Excluded - research and development process
Users-producers seminars		Included as Qualitative methods of user interaction
Usability tests		Excluded - research and development process
Users feedbacks		Excluded - research and development process
Clinical trials		Excluded - research and development process
Choice-based conjoint analysis (Discrete choice modelling)	D**iscrete choice experim**ents	Included as Discrete choice experiments

Horizon-scanning		Excluded – horizon scanning is a process using methods such as literature review and stakeholder engagement to identify emerging technologies. It may be useful in DF-HTA to identify barriers to diffusion and relevant comparators for economic evaluation. Discussed under literature review in Section 5.3.3.3.
Preliminary market research		Excluded - other commercial activities
Bench studies		Excluded - research and development process
Quantitative		
Headroom analysis		Included under Cost-effectiveness analysis
Cost-effectiveness analysis	**Cost-effectiveness analysis**	Included
Probabilistic sensitivity analysis		Included under Cost-effectiveness analysis
Potential years of life lost		Included under Cost effectiveness analysis
Cost-benefit analysis		Included under Cost-effectiveness analysis
Cost-utility analysis		Included under Cost-effectiveness analysis
Opportunity costs (used as indicators to which relative weights are assigned)		Excluded as relevance not clear
Road-mapping process (Multi-path Mapping)		Excluded - research and development process
Scenarios building		Included under Cost-effectiveness analysis
Return on investment		Excluded - other commercial activities. Discussed under cost effectiveness analysis as headroom estimates can inform return on investment calculations
Technological forecasting based on epidemiological data		Excluded - other commercial activities
Rudimental analysis of costs		Included as cost-minimisation analysis under Cost-effectiveness analysis
Multi-criteria Decision Analysis (Analytic Hierarchy Process)	**Multi-criteria Decision Analysis**	Included
Expected value of perfect information		Included under Value of information analysis
Bayesian modelling/statistics (data pooling, random effects analysis)		Included under Cost-effectiveness analysis
Probabilistic risk analysis		Excluded - research and development process
Real options analysis		Excluded - other commercial activities. Discussed under Cost effectiveness analysis as headroom can be combined with a form of real options analysis.
Best-worst scaling		Included as Discrete choice experiments
Decision tree analysis		Excluded as forms part of other modelling approaches discussed in Cost effectiveness analysis
Methods from Ijzerman and Steuten (2011)		

Payback from research analysis		Excluded - other commercial activities
Strategic business case		Excluded as umbrella term
Health impact assessment		Excluded as umbrella term
Multi-criteria Decision Analysis (Analytic Hierarchy Process)		Excluded as duplicate
Choice-based preference methods (discrete choice experiments and conjoint analysis)		Included as Discrete Choice Experiments
Real options analysis		Excluded as duplicate
Health economic modelling		Included under Cost-effectiveness analysis
Horizon-scanning systems		Excluded as duplicate
Clinical trial simulation		Excluded - research and development process. Discussed under Value of Information.
Value of information analysis	**Value of Information Analysis**	Included
Methods from Redekop and Mikudina (2013)		
Early health economic modelling		Included under Cost-effectiveness analysis
Clinical trial simulation		Excluded as duplicate
Multi-criteria Decision Analysis		Excluded as duplicate
Headroom		Included under Cost-effectiveness analysis
Bayesian analytical framework		Included under Cost-effectiveness analysis
Value of Information analysis		Excluded as duplicate
Methods from Miller (2005)		
Clinical Trial Simulation		Excluded as duplicate
Option Pricing		Excluded as synonymous with real option analysis
Investment Appraisal		Excluded as umbrella term
Threshold analysis		Included under Cost-effectiveness analysis
Value of information analysis		Excluded as duplicate
Methods from Hartz and John (2008)		
Early health economic modelling		Excluded as duplicate
The Bayesian Analytical Framework		Excluded as duplicate
Value of information analysis		Excluded as duplicate
Clinical trial simulation		Excluded as duplicate
Methods from Bartelmes (2009)		
Analytic Hierarchy Process		Included as Multi-criteria Decision Analysis
Stated-preference methods		Included as Discrete choice experiments
Expert systems		Excluded - research and development process

Fuzzy logic		Excluded - other commercial activities
Bayesian methods		Excluded as duplicate
Decision analytic models (e.g. Markov models)		Excluded as duplicate
Pharmacokinetic and pharmacodynamic modelling		Excluded - research and development process
User-centred design		Excluded - research and development process
Failure and reliability analysis		Excluded - research and development process
Real-options analysis		Excluded as duplicate
Pre-protocol research		Excluded - research and development process
Tracker-trials		Excluded - research and development process
Constructive Technology Assessment		Excluded as umbrella term
Iterative economic evaluations		Included under Cost-effectiveness analysis
Evaluation frameworks for information technologies		Excluded as not development-focused
Methods from Graziadio (2020)		
Articulating value propositions		Included as outcome in Qualitative methods of user interaction
Care pathway analysis	**Care pathway analysis**	Included
Clinical validity studies		Excluded as research and development
Clinical utility studies		Excluded as research and development
Cost-effectiveness analysis		Included
Cost-consequences analysis		Included under Cost-effectiveness analysis
Budget impact analysis		Included under Cost-effectiveness analysis

Appendix 7 - Examples of economic evaluation undertaken in development-focused HTA

First author (date) Jurisdiction	Technology	Decisions to be informed	Research and development stage	Method and model	Estimands	Deals with uncertainty	Data for decision problem	Data for costs	Data for effects
Brandes (2015) Germany	Vascular closure device in cardiac surgery	Reimbursement/regulatory strategy Technology design	Hypothetical device	CEA - Decision tree 24-hour time horizon	Value-based price per complication avoided. Cost savings	OWSA Scenario analysis	Literature	Public sources	Literature
Braz (2013) Netherlands and Portugal	Photo-acoustic imaging in breast cancer	Reimbursement/regulatory strategy	Phase I trials underway	CUA – Markov model 9-29 years time horizon	ICER	Scenario analysis	Building on previous work by the same group	Public sources	Literature and expert elicitation for clinical effectiveness
Buisman (2016a) Netherlands	Diagnostic strategies in rheumatoid arthritis	Reimbursement/ regulatory strategy Go/no-go - ongoing investment	Tests under development	CUA - Decision tree – 12 month time horizon Individual level state transition model – 4 year time horizon	Headroom ICER	OWSA PSA Scenario analysis	TRACER project Dutch guidelines	Cohort study and public sources	Effects from cohort study and assumptions, test performance from developers
Buisman (2016b) Netherlands	Diagnostic strategies in rheumatoid arthritis	Reimbursement/ regulatory strategy Technology design Go/no-go - ongoing investment	Tests under development	CEA - Decision tree Individual level state transition model 5 year time horizon	Headroom Test characteristics required	OWSA PSA	TRACER project	Cohort study and public sources	Effects from cohort study and assumptions, test performance from developers

First author (date) Jurisdiction	Technology	Decisions to be informed	Research and development stage	Method and model	Estimands	Deals with uncertainty	Data for decision problem	Data for costs	Data for effects
Cao (2013) Not stated	Home monitoring using point of care test in cardiovascular disease	Reimbursement/ regulatory strategy Go/no-go - ongoing investment	Device in development	CEA - Conceptual models Continuous-time Markov Model	Headroom	PSA	Expert elicitation	Literature	Cohort/ Expert elicitation for transition probabilities
Carr (2013) Not stated	Tissue engineering in urinary stress incontinence	Evidence generation strategy Go/no-go - ongoing investment	Pre Phase I trials	CUA – Decision tree 5 year time horizon	Net health benefit	Simulation for uncertainty in costs. Point estimates for other parameters.	Literature	Literature and simulation model	Literature and simulation model
Chapman (2013) United Kingdom	Multiple technologies (short retrospective case studies), COPD and leg ulcers (prospective studies)	Reimbursement/ regulatory strategy Go/no-go - ongoing investment	Concept stage for prospective studies. For the retrospective studies author put herself in the position of assessing at an early stage in development.	CUA - Equation	Headroom Maximum price Cost savings Optimum pathway position Maximum development costs	Scenario analysis	Guidelines, expert opinion, literature, developers	Literature, developers, public sources	Utilities from literature or derived
Craven (2011) United Kingdom	Electrical stimulation for diabetic foot ulcer	Reimbursement/ regulatory strategy	Not stated	CUA - Markov model – 20 year time horizon	Incremental Cost Effectiveness Ratio (ICER)	Scenario analysis	Literature	Public sources	Literature and developer information, assumptions
Davey (2011) United Kingdom	Stent for peripheral arterial disease and tissue engineering for cancer of the bladder	Reimbursement/ regulatory strategy	Early phase of development	CUA - Equation	Ranked projects	Not applicable	Expert opinion	MCDA informed by headroom analysis	Expert opinion

First author (date) Jurisdiction	Technology	Decisions to be informed	Research and development stage	Method and model	Estimands	Deals with uncertainty	Data for decision problem	Data for costs	Data for effects
De Graaf (2015) Netherlands	Multiple biomarkers in Type 2 diabetes	Portfolio management	Prior to translational research	CUA - Equation MCDA	Headroom	Not stated	Expert opinion	Literature Expert elicitation	Literature Expert elicitation
De Graaf (2018) Netherlands	Biomarkers to predict CVD risk in patients with Type 2 diabetes	Reimbursement/ regulatory strategy Portfolio management	Prior to translational research	CUA - Equation	Clinical impact (number of treatments withheld, CVD events) Headroom	OWSA	Expert opinion	Public sources	Cohort study
Gaultney (2011) Netherlands	Companion diagnostic in Chronic Myeloid Leukemia	Reimbursement/ regulatory strategy Evidence generation strategy	Prototype	CUA - Decision tree 2 year time horizon	Costs Progression Free Life Years QALYs	OWSA and two-way sensitivity analysis Scenario analysis	Dutch guidelines	Insurance board	Literature review (clinical trials) and expert opinion
Hjelmgren (2006) Sweden	Dopamine cell replacement in Parkinson's disease	Evidence generation strategy Technology design	Post small clinical trial	CUA – Simulation model 20 year time horizon	ICER	OWSA Scenario analysis	Research team	Previous study	Progression from cohort study Effectiveness from two small trials
Hummel (2012) Netherlands	Non-fusion surgery in adolescent idiopathic scoliosis	Evidence generation strategy Go/no-go - ongoing investment	Innovative form of surgery	CUA – Decision tree	Quality of life Complications Costs	OWSA	Not stated	Insurance data and micro-costing	Literature for disease progression and current treatment. MCDA with experts for new treatment
Huygens (2016) Netherlands	Tissue engineered heart valve in CVD	Technology design for cost-effectiveness	Concept	CEA – Conceptual model	Conceptual model only	Not applicable	Clinical guidelines and Delphi panel of experts	Not applicable	Not applicable

First author (date) Jurisdiction	Technology	Decisions to be informed	Research and development stage	Method and model	Estimands	Deals with uncertainty	Data for decision problem	Data for costs	Data for effects
Kenter (2015) Netherlands	Circulating cancer cell trap (CTC trap) in prostate cancer	Reimbursement/ regulatory strategy	Prototype	CUA - Decision tree Various Markov models – unclear duration	ICER Costs QALYs	PSA	Literature search, clinical guidelines and expert opinion	Literature and assumptions	Literature and assumptions
Khoudigian-Sinani (2017) Canada	Diagnostic test (Straticyte) in oral cancer	Reimbursement/ regulatory strategy	Pre-market assessment	CEA/CUA - Conceptual model Decision tree – 5 year time horizon	Cancer cases prevented Costs per cancer case avoided ICER	OWSA PSA Scenario analysis	Experts and systematic review	Literature, experts, developers	Belief elicitation, literature and observational data for effects and utilities
Kip (2016) Netherlands	Biomarker test in suspected heart attack	Reimbursement/ regulatory strategy Technology design	Pre-clinical application	CEA – Decision tree 24 hour time horizon	Percentage of patients correctly discharged Direct hospital costs	OWSA PSA Scenario analysis	Clinical guidelines and expert opinion	Dutch reimbursement data and published sources	Expert elicitation based on assumed test performance
Kluytmans (2019) Netherlands	Surgical instrument in meniscus surgery	Go/no-go - ongoing investment	In development	CUA – Decision Tree	Headroom Threshold performance at fixed price level	OWSA	Expert opinion	Assumptions Prices from single medical centre	Expert opinion Literature Public sources
Knuttell (2017) Netherlands	Magnetic resonance imaging (MRI) guided ablation in breast cancer	Technology design for cost requirements	After small clinical studies	CMA – Decision tree over duration of acute treatment	Costs of new treatment	OWSA	Expert opinion	MRI imaging costs as proxy for new procedure, micro-costing for timing. Literature for other costs	Registry data and small clinical studies

First author (date) Jurisdiction	Technology	Decisions to be informed	Research and development stage	Method and model	Estimands	Deals with uncertainty	Data for decision problem	Data for costs	Data for effects
Koerber (2013) Germany	Regenerative medicine in knee cartilage repair	Reimbursement/ regulatory strategy Evidence generation strategy Go/no-go - ongoing investment	Phase 0-II clinical trials	CUA - Decision tree 2 year time horizon	Costs Quality Adjusted Life Years	Mentions scenario analysis and sensitivity analysis but none presented	Guidelines and interviews with clinicians	Literature and public sources	Literature
Kolominsky-Rabas (2015) Germany	Mobile stroke units	Go/no-go - ongoing investment Technology design	Prototype available	CEA – Hybrid simulation	Time to treatment Proportion of patients receiving thrombectomy	Built into simulation	Not stated	Not stated	Not stated
Latimer (2011) United Kingdom	Neck collar in motor neurone disease	Go/no-go - ongoing investment	Concept	CUA – Decision tree Markov model 4 year time horizon	Threshold cost for cost-effectiveness	Scenario analysis	Clinical audit at single hospital	Patient group for existing options	Patient group for disease progression Elicitation for utilities
Liu (2013) United States	Home management point of care test in chronic obstructive pulmonary disease	Technology design Go/no-go - ongoing investment	Hypothetical home-management technology	CUA - Markov model 3-12 year time horizon	Patient lifetime savings Acceptable price for technology Cost boundaries for different risk groups	PSA Scenario analysis	Experts opinion Observational data Developers	Medicare reimbursement schedules	Experts opinion Observational data Literature
Luime (2016) Netherlands	Diagnostics in rheumatoid arthritis	Reimbursement/ regulatory strategy Technology design	Tests in development	CUA – Decision tree 1 year post-diagnosis	Headroom ICER	OWSA PSA	Guidelines and current study	Test costs from developer Treatment cost from observational study	Clinical studies for test performance Registry data Observational data

First author (date) Jurisdiction	Technology	Decisions to be informed	Research and development stage	Method and model	Estimands	Deals with uncertainty	Data for decision problem	Data for costs	Data for effects
Markiewicz (2016) Netherlands	Multiple devices (1 therapeutic, 5 diagnostic)	Reimbursement/ regulatory strategy Technology design Portfolio management	Real time to support decision making in translational research portfolio	CUA - Equation	Headroom Return on investment	No	Developer Literature Clinical expert opinion	Developer Literature Clinical expert opinion	Developer Literature Clinical expert opinion
McAteer (2007) United Kingdom	Tissue engineering in urogenital medicine	Go/no-go - ongoing investment	Proposed tissue applications	CUA - Equation	Headroom	No	Expert opinion and literature searches, technical updates from developers	Expert opinion	Time trade-off among experts
Postmus (2012) Netherlands	Biomarker for screening in type 2 diabetes	Technology design – required test performance	Hypothetical biomarker and intervention in type 2 diabetes	CUA - Discrete time Markov model	Headroom, Quality adjusted life years Net Reclassification Index	OWSA	Systematic review	Literature and expert opinion	Literature and expert opinion
Rejon-Parilla (2014) Netherlands	Biomarkers for guided dosing in schizophrenia	Technology design - required test performance	Biomarker has been associated with differential response but not used to guide dosing	CUA - Markov model 2 year time horizon	ICER	OWSA Scenario analysis	Literature	Literature	Literature
Singh (2010) United Kingdom	Regenerative medicine in hearing loss	Go/no-go - ongoing investment	Hypothetical technology	CUA - Equations	Headroom Return on investment	Key parameters and uncertainty around them dealt with qualitatively.	Literature and developer information	Literature and developer information	Literature, developer information and assumptions for utility gain
Tan (2015) Not stated	Corneal endothelial transplantation	Go/no-go - ongoing investment	Pre-phase I trials	CMA – unclear	Costs of production	OWSA PSA	Unclear	Unclear	Unclear

First author (date) Jurisdiction	Technology	Decisions to be informed	Research and development stage	Method and model	Estimands	Deals with uncertainty	Data for decision problem	Data for costs	Data for effects
Vallejo-Torres (2011) United Kingdom	Absorbable pins in hallux valgus	Evidence generation strategy Go/no-go - ongoing investment	Retrospective analysis mimicking concept stage	CUA – Decision tree 12 month time horizon	Headroom Cost saving QALY ICER EVPI EVPPI	OWSA PSA	Observational data Randomised controlled trial	Public sources	Utilities from expert elicitation Current effectiveness from literature Assumptions for new technology
van Nimvegen (2017) United Kingdom	Diagnosis of complex paediatric neurology	Go/no-go - ongoing investment	Hypothetical technology	CEA - Equation	Headroom	Scenario analysis	Literature	Literature	Literature
van de Wetering (2012) Netherlands	Lab on a chip - point of care device for patients with heart failure	Reimbursement/ regulatory strategy Evidence generation strategy Technology design Go/no-go - ongoing investment	Technology currently in development	CUA - Markov model 5 year time horizon	Headroom ICER	PSA Scenario analysis	Literature, assumptions and information from developer	Literature	Literature and interviews
Vilsboll (2018) Denmark	Cell-based therapy for female stress urinary incontinence	Evidence generation strategy Technology design Go/no-go - ongoing investment	Clinical trials ongoing	CUA – Decision tree 1 year time horizon	Headroom ICER	OWSA	Guidelines	Public sources and expert opinion	Small clinical trial Expert opinion Observational data Literature

CCA – cost-consequence analysis, CEA – cost-effectiveness analysis, CMA – cost-minimisation analysis, COPD – chronic obstructive pulmonary disease, CTC trap – circulating cancer-cell trap, CUA – cost-utility analysis, CVD – cardio-vascular disease, EVPI – Expected value of perfect information, EVPPI – expected value of partial parameter information, ICER – incremental cost-effectiveness ratio, MCDA – multi-criteria decision analysis, MRI – magnetic resonance imaging, OWSA – one-way sensitivity analysis, PSA – probabilistic sensitivity analysis, QALY – Quality adjusted life year,

Appendix 8 - Studies where expert elicitation has been applied in development-focused and early HTA

Development-focused HTA			
First author (date)	**Method/participants**	**Parameters elicited**	**Purpose**
Davey (2011)	Meetings Unspecified number of vascular surgeons and interventional radiologists	Clinical treatment routes, disease context, current treatments and current technology characteristics	To identify the drivers for clinical success
Breteler (2012)	Interviews 18 neurologists	Likely test performance, position in diagnostic pathway, acceptance by patients, technical success	To populate a real options analysis
Weiser (2013)	Self-administered questionnaire 38 rheumatologists	Characteristics required of new imaging modality in Rheumatoid Arthritis	To direct the development of photo-acoustic imaging in rheumatoid arthritis
Haakma (2014)	Bayesian belief elicitation 18 radiologists	True positive and true negative rates for new imaging technology	To populate a cost-effectiveness model
Huygens (2016)	Delphi panel 10 cardiothoracic surgeons, cardiologists and a biomedical scientist	Conceptual model for cost-effectiveness analysis of Tissue engineered heart valve in cardiac surgery	First step to cost-effectiveness model that will investigate the performance and costs requirements for a new heart valve technology
Kip (2016)	Expert elicitation by questionnaire. 10 cardiologists.	The effect of the new biomarker on hospital discharge rates and interventions performed	To populate cost-effectiveness models to guide further investment decisions

Early HTA			
First author (date)	**Method/participants**	**Parameters elicited**	**Purpose**
Girling (2007)	Bayesian prior distributions elicited in group setting 5 cardiologists	Survival prospects of patients treated with second generation device	To investigate the survival benefit required for technology to be cost-effective
Leal (2007)	Expert elicitation using computer-based tool pooled using linear opinion pooling 7 experts comprising cardiologists, clinical geneticists and laboratory scientists	Probability distributions for test performance, population at risk, detection by cardiology services, effectiveness of interventions	To populate an early health economic model to inform decision making for new genetic testing in hypertrophic cardiomyopathy
Oestergaard (2010)	Delphi panel 12 authors of clinical studies on the genotyping of interest	Response rates to treatments for depression following genetic testing	To predict the extent to which pretesting for 5-HTTLPR would improve health outcomes
Terjesen (2017)	Delphi panel by two round questionnaire 8 clinicians and/or researchers	Probability of cross-contamination from use of existing technology	To confirm a core aspect of the value proposition/clinical need

5 -HTTLPR - serotonin-transporter-linked polymorphic region

Appendix 9 - Studies where multi-criteria decision analysis has been applied in development-focused and early HTA

Development-focused HTA			
First author (date)	**Method/participants**	**Criteria identified**	**Purpose**
Hummel (2000)	Team Expert Choice – a group decision support system based on AHP Cardiologist, 6 engineers (chemical, mechanical and electrical), surgeon, and veterinarian. Group supported by independent facilitator	Performance Safety Ease of use Applicability	To guide development and diffusion of the Pulsatile Catheter Pump
Hilgerink (2011)	AHP Panel composed of medical, technical and management experts	Costs Diagnostic performance Patient comfort Risks	To assess the added value of photo-acoustic imaging in the diagnosis of breast cancer
Hummel (2012)	AHP Panel of biomedical engineers and orthopaedic surgeons	Costs (materials and treatment) Effectiveness (quality of life, pain, back function, self-esteem, medical and technical complications)	To predict the health economic performance of new non-fusion surgery in adolescent idiopathic scoliosis
Wissing (2012)	AHP Patients Endocrinologists	Clinical effectiveness Long term patient safety Complications	To prospectively evaluate if and when, the bioartificial pancreas becomes an alternative to transplantation methods in diabetes type 1
Koning (2012)	MCDA (using a basic version of AHP) Microbiology physician, laboratory manager, head of the department, hospital management	Clinical performance (accuracy, time to diagnosis, range of micro-organisms) Cost of ownership Impact on workflow	To populate performance criteria for an early health economic evaluation of an electronic nose technology
De Graaf (2015)	MCDA with Stochastic multi-criteria acceptability analysis adapted to incorporate ordinal data. Participants were researchers from the specific project	Reduction in downstream costs Added quality adjusted survival Cost Feasibility of treat-all Competition Ease of implementation	To support priority setting within a research team looking at biomarkers in Diabetes Type 2 prevention. Prior to research being undertaken

AHP – Analytic Hierarchy Process, MCDA – Multi-criteria Decision Analysis

Early HTA			
First author (date)	**Method/participants**	**Criteria identified**	**Purpose**
Lambooij (2013)	AHP Stakeholder groups of nurses, patients, physicians, managers, health-care insurers and policy makers	Improvement in efficiency Health gains Satisfaction with care process Investments required	To explore differences in stakeholder preferences between nine information technology innovations in hospital care

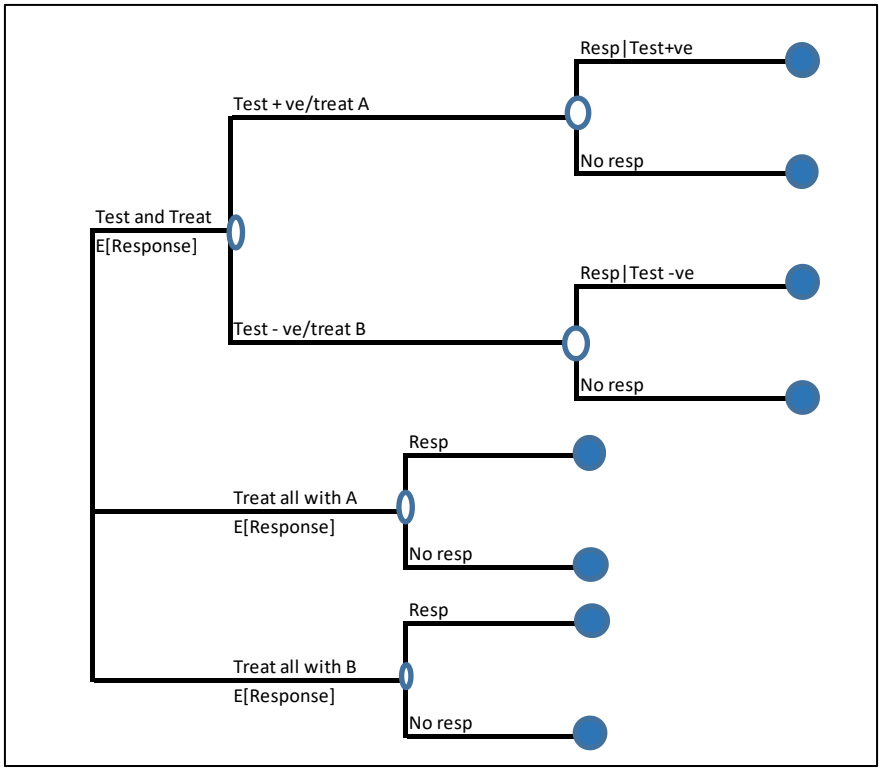
AHP – Analytic Hierarchy Process, MCDA – Multi-criteria Decision Analysis

Appendix 10 – Test and treatment threshold plot - technical detail

The probability of response to treatment A is pA and the probability of response is pB. A test is available that will predict response to treatment A with sensitivity (Sens) and specificity (Spec). The probability of response to treatment B is independent of the test result.

The decision problem is illustrated by Figure AP10-1.

Figure AP10-1: Generic decision model



The difference in expected response between the test and treat strategy and the optimal treatment strategy is given by:

$$\max(pA, pB) - pA.pB + Sens.pA - Sens.pA.pB + Spec.pB - Spec.pA.pB \quad \text{Eqn. 1}$$

The threshold sensitivity at which the expected response from the test and treat strategy is equal to the treatment strategy (when A is the optimal treatment) is given by:

$$sens = \frac{(1-pB+Spec.pB-(Spec.pB))/pA)}{(1-pB)} \quad \text{Eqn. 2}$$

The threshold sensitivity at which the expected response from the test and treat strategy is equal to the treatment strategy (when B is the optimal treatment) is given by:

$$sens = \frac{pA + Spec - Spec.pA - 1}{pA - \frac{pA}{pB}} \qquad \text{Eqn. 3}$$

This can be used to establish thresholds for the sensitivity and specificity at which the test and treat strategy will be superior to the optimal treat only strategy. This is a function of the odds ratio for treatment A vs treatment B.