Incorporating technology diffusion estimates in health economic methods

Application in a preterm birth screening case study

Sabine Elisabeth Grimm, MSc, BA

Thesis for the degree of Doctor of Philosophy

Supervised by Dr John W Stevens and Professor Simon Dixon

Health Economics and Decision Science, School of Health and Related Research, University of Sheffield

Submitted in May 2016

ABSTRACT

Low implementation of cost-effective health technologies results in inefficient use of resources in a health system. Despite this, estimates of implementation or diffusion are not routine components of analyses performed within health technology assessments (HTA), potentially due to a lack of a) methods to obtain diffusion estimates and b) understanding of the impact of diffusion estimates on health economic outcomes. This thesis contributes a) a method to estimate health technology diffusion prior to HTA and b) a modelling framework that assesses the potential impact of diffusion estimates on cost-effectiveness and expected value of information and implementation (EVII) analysis using modelling, qualitative and elicitation methods. These were illustrated in a preterm birth (PTB) screening case study.

The modelling framework included extensions to an existing EVII model to make it dynamic and allow research to affect implementation; and the development of a dynamic cost-effectiveness analysis (DCEA) model that reflects price changes precipitated by diffusion and hence, the reimbursement decision. Drivers of diffusion were identified for the case study technology, aiding the design of implementation strategies. The developed method for predicting diffusion requires transformation of elicited expert beliefs to inform an existing diffusion model. Application in the PTB screening model showed that the dynamic EVII method can 1.) help more accurately assess the losses the health care payer incurs when there is decision uncertainty and low implementation and 2.) provide more realistic assessments of implementation strategies and evidence generation schemes. The applied DCEA model showed that changes in price triggered by technology diffusion significantly affect cost-effectiveness results. The method for predicting health technology diffusion and the EVII and DCEA frameworks are foreseen to be relevant in the context of HTAs of medical devices, diagnostics and drugs; particularly when there is low implementation or there is potential for future price changes conditional on diffusion.

ACKNOWLEDGEMENTS

I wish to thank HEDS and EEPRU for funding this research study. The opportunity to pursue a PhD will always be greatly appreciated. I am grateful to my supervisors John Stevens and Simon Dixon for making this a better thesis with their contributions and advice throughout this period, and for everything I learnt from them.

I extend my gratitude to all experts that were interviewed, with special thanks to those who also volunteered for participating in the elicitation study. I also thank Dilly Anumba for sharing information on the case study technology and Sylvia Bates for volunteering to pilot the elicitation study.

Many people in HEDS have supported me in many ways and I am grateful to them. I would also like to thank the other ScHARR PhD candidates who shared the joys and the pains of the process and provided invaluable advice, especially Milad Karimi, Sarah Smith, Penny Breeze, and Kiera Bartlett.

My parents Ralph and Beate have shown infinite patience with my long road of education and I am very grateful to them for their support, as well as to my late grandmother Helga who was and remains an inspiration to me. I also thank my sister Marietta and my brother Jan for their encouragement throughout this process. I am indebted to my partner Bernard, whose sense of humour and untiring support enabled me to finish this thesis. Thanks also to our little Phileas, whose infectious giggle brightened up the second half of this study.

DISSEMINATION

I have submitted parts of this work to the following journals:

- Medical Decision Making: Grimm, S, Dixon, S, and Stevens, JW; 'Assessing the expected value of research studies in reducing uncertainty and improving implementation dynamics'; revise & resubmit; revisions currently under review
- Value in Health: Grimm, S, Dixon, S, and Stevens, JW; 'When future change matters: modelling future price and diffusion in health technology assessments of medical devices'; revise & resubmit; revisions currently under review

I have prepared another manuscript that I am about to submit to the following journal:

- International Journal of Forecasting: Grimm, S, Stevens, JW, Dixon, S; 'A new pre-launch forecasting method using elicitation of expert beliefs and an adaptation of the Bass model of new product growth'

Parts of this work are available online within the HEDS discussion paper series:

- Grimm, S, Dixon, S, and Stevens, JW (2015); 'Assessing the expected value of research in resolving uncertainty and improving implementation'; https://www.sheffield.ac.uk/scharr/sections/heds/discussion-papers/15-03-1.516976
- Grimm, S, Dixon, S, and Stevens, JW (2013); 'Are we over-estimating the value of further research?'; https://www.sheffield.ac.uk/scharr/sections/heds/discussion-papers/1315-1.329384

A part of this thesis has informed a report:

- Essat, M., Faria, R., Gomersall, T., Grimm, S., Keetharuth, A., Walker, S., Dixon, S., Palmer, S., Sculpher, M. (2013). 'Getting cost-effective technologies into practice: the value of implementation. Initial scoping review for phase 1.' Policy Research Unit in Economic Evaluation of Health and Care Interventions (EEPRU) 2014; Research Report 023; Available online:

www.eepru.org.uk/Value%20of%20implementation%20-%20Report%20on%20reviews.pdf

I also presented (and am going to present) parts of this work at the following national and international conferences and seminars:

- SMDM London, UK, June 2016; 'Assessing the expected value of research resolving uncertainty and improving implementation' (poster presentation)
- HEDS seminar series, University of Sheffield, UK, February 2015; 'Estimating the impact of health technology implementation on cost-effectiveness and value of information analysis' (oral presentation)
- HTAi Washington, DC, USA, June 2014; 'Evaluating implementation strategies – modelling the link between information and implementation dynamics using data obtained from elicitation of expert opinions' (oral presentation)
- Workshop on Technology Assessment of Medical Devices, Czech Technology University, Prague, Czech Republic, June 2013; 'Economic evaluation of electrical impedance spectroscopy for use in pre-term birth screening: incorporating uptake estimates' (oral presentation)
- HESG Meeting, Exeter, UK, January 2013; 'Do we care about the future? Why implementation levels should be estimated in early-stage economic evaluation' (paper presented by another participant with subsequent author clarifications)
- Health Economics seminar series, University of Birmingham, UK, December 2012; 'Do we care about the future? Why implementation levels should be estimated in early-stage economic evaluation' (oral presentation)

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

BPEG	British Pacing and Electrophysiology Group
CEA	Cost-effectiveness analysis
CEAC	Cost-effectiveness acceptability curve
CI	Credible interval
CL	Cervical length
CQUIN	Commissioning for Quality and Innovation
CTA	Constructive technology assessment
DGH	District General Hospital
DH	Department of Health
EIS	Electrical impedance spectroscopy
ePACT	Electronic prescribing analysis tool
ERNIE	Evaluation and Review of NICE Implementation Evidence
EVI	Expected value of information
EVII	Expected value of information and implementation
EVIM	Expected value of implementation
EVP	Expected value of perfection
EVPI	Expected value of perfect information
EVPIM	Expected value of perfect implementation
EVR	Expected value of research
EVSI	Expected value of sample information
EVSIM	Expected value of specific implementation measures
GAM	Generalised additive model
GBM	Generalised Bass model
GBP	British pound sterling
GP	General Practitioner
HES	Hospital episode statistics
HTA	Health technology assessment
ICER	Incremental cost-effectiveness ratio
IM	Implementation measure
LR	Likelihood ratio
MRI	Magnetic resonance imaging
NB	Net benefit
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
OAG	Open angle glaucoma
PCA	Prescription cost analysis
PEVI	Population expected value of information

PSA	Probabilistic sensitivity analysis
PSS	Personal social services
PSSRU	Personal Social Services Research Unit
PTB	Preterm birth
QALY	Quality-adjusted life year
QOF	Quality and Outcomes Framework
RwR	Recommendation with research
SD	Standard deviation
UK	United Kingdom
UTH	University Teaching Hospital

CHAPTER 1. INTRODUCTION: WHY DIFFUSION ESTIMATES ARE IMPORTANT TO HEALTH TECHNOLOGY ASSESSMENT

1.1 Chapter outline

The aim of this chapter is to introduce the research undertaken for this thesis. In Section 1.2, I explain the background to this thesis. Section 1.3 presents the rationale for and the hypotheses to be tested in this thesis, as well as an introduction to the case study on electrical impedance spectroscopy for use in preterm birth screening and my philosophical position. In Section 1.4, I describe my research aim and objectives and Section 1.5 presents the structure of the thesis.

1.2 Background on technology implementation and health technology assessment

1.2.1 Definitions of key concepts

First, I wish to highlight that this thesis adopts the perspective of the health care payer, and, to make the research questions in this thesis specific and ensure its feasibility, I focus on one jurisdiction, England. The National Institute for Health and Clinical Excellence (NICE) is the decision-making body for reimbursement of health technologies in England.

The topic of this thesis focusses on health technology uptake. The term 'health technology' is defined as any product or activity that is 'used to promote health, prevent and treat disease and improve rehabilitation and long term care' (NIHR, 2012) and can thus include medicines, medical devices, procedures and screening technologies, amongst others. 'Uptake' describes the number of units of a technology in use as a share of the maximum number of desirable units, typically at one point in time. The health economic literature often uses the term 'implementation' to describe uptake (Fenwick et al., 2008, Department of Health, 2011). I use the terms 'uptake' and 'implementation' interchangeably throughout this thesis, although different interpretations have been observed in different contexts

(Greenhalgh et al., 2004). The terms ‘adoptions’ or ‘utilisation’ are also commonly used to describe the units in use of a technology. As it is defined above, uptake is a relative measure. However, both terms of uptake and implementation have been used as absolute measures of the number of adoptions, which would ultimately be compared with the maximum number of attainable adoptions. Technology uptake can confusingly also be understood as a dynamic process. To make a clear distinction, I refer to this dynamic process as diffusion, which is a term that has been used to describe the process of technology uptake, for instance by Rogers in his book ‘Diffusion of Innovations’ (Rogers, 2003).

The definition of the uptake concept is further complicated by the different levels of uptake that can occur in health technologies. To make this clearer, I use three examples: 1. the adoption levels of a medical device could entail a) the number of hospitals purchasing the device; b) the number of clinicians using the device, or alternatively, the number of patients being offered the device; and c) the number of patients ultimately receiving the device, taking patient acceptance into account. 2. The adoption process of a drug mainly entails the number of physicians prescribing the drug and the number of patients accepting it. 3. For screening technologies, the only level of uptake is often the patient level, or that of the general public. In this thesis, I use a medical device as a case study example. To simplify the illustration of the concepts developed in this thesis in the case study technology, I focus on a) the number of hospitals purchasing the device.

In this thesis, I use the term ‘implementation strategies’ to refer to any initiatives and programmes that are aimed at increasing the uptake of health technologies (Essat et al., 2013) in the England NHS. Other terms commonly used in the literature are ‘implementation initiatives’ and ‘implementation measures’ (Essat et al., 2013), and I use these interchangeably.

1.2.2 Health technology assessment and health economic methods

With a fixed budget for health care spending in England, resource allocation decisions are typically made using coherent frameworks for decision-making that are based on health economic principles (Drummond et al., 2005). These principles entail the maximisation of benefits to patients in the NHS within a given budget. Welfarism demands that overall welfare, measured in terms of individuals’ evaluations of their utility associated with different states, be maximised subject to budget constraints. Assigning monetary value to utilities, cost-benefit analyses are used to implement welfarism in practice (Morris et al., 2007). There are, however, various limitations associated with welfare economics used in

health care. These include that welfarism is firmly based in consumer choice theory that makes assumptions such as consumer rationality, amongst others, which cannot be guaranteed in the health care setting in which market failure is commonly assumed to occur (Morris et al., 2007). Furthermore, individual utilities are not deemed a good measure of well-being (Morris et al., 2007), as the concept of utility in its classical sense relates to utility arising from consumption of goods.

The most commonly used and recommended approach to resource allocation decisions in England is therefore based on so-called 'extra-welfarism', a framework that permits measures other than utilities, for example measures of health, in the social welfare function (Morris et al., 2007). With it comes an alternative to cost-benefit analysis, the so-called cost-effectiveness analysis that allows measuring benefits in non-monetary terms (Morris et al., 2007). This compares costs of decision options with the health gain that can be achieved, and if this is expressed in terms of the quality-adjusted life year (QALY), which is typically used in decision-making in England (NICE, 2013), then the correct term for this type of analysis is cost-utility analysis (Drummond et al., 2005). I will continue to use the term cost-effectiveness analysis, as this appears to be more widely used, certainly in the context of NICE decision-making (NICE, 2013). The decision of whether a new health technology is a good use of resources compared to other options, that is, whether it is cost-effective, is made using a decision rule. This decision rule entails two components: first, the incremental cost per QALY, called the incremental cost-effectiveness ratio (ICER) of a health technology over an alternative is calculated. Second, the ICER is compared with a pre-specified threshold that represents the ceiling value that the health care payer would pay for a QALY given the budget constraints in the health system (Eckermann and Pekarsky, 2014). The threshold for most drugs is set to a range of £20,000 to £30,000 per QALY in England (NICE, 2013). Thresholds for other health technologies may be different, for instance the threshold for medical devices is £0 per QALY, expecting these to be cost-saving compared to current practice (NICE, 2012).

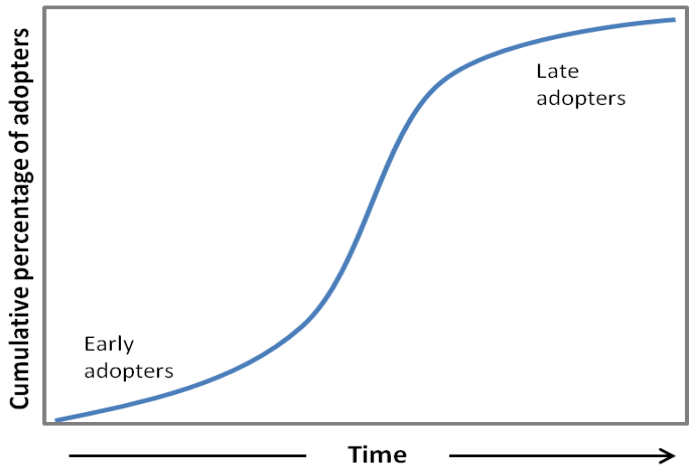
The aim of cost-effectiveness analyses within health technology assessments is to provide estimates of the ICER associated with a new health technology compared with existing alternatives (NICE, 2013). The process for estimating costs and QALYs associated with different health technologies can entail complex mathematical modelling methods, the synthesis of evidence from different sources and simulation methods to consider the uncertainty associated with the input parameters (Briggs et al., 2006). Value of information methods, while not routinely used in the process of health technology assessment (HTA) (NICE, 2013), can inform the decision-maker on the opportunity loss associated with uncertainty in a decision and, beyond that, can inform research design decisions (Briggs et

al., 2006, Claxton and Posnett, 1996, Brennan and Kharroubi, 2007, Strong et al., 2015, Eckermann and Willan, 2007). For this, the value of information is scaled up over the affected patient population and accrued over future time periods up to the decision relevance horizon, that is, the time until which a decision remains relevant, for example until there is a new technology class replacing the current one. The decision relevance horizon has in the past been most relevant to value of information analysis (Philips et al., 2008) because future periods are not generally modelled in cost-effectiveness analyses (NICE, 2013). Philips et al. (2008) found that the nature and magnitude of change, for example in price, evidence and / or competition affected the length of the decision relevance horizon.

1.2.3 The dynamic nature of technology implementation

Uptake of new technologies typically follows a dynamic process (Rogers, 2003). This process has been much described in the marketing literature as that of new product growth or diffusion (Meade and Islam, 2006). A theoretical framework was developed by Rogers (2003) in the first edition of his book 'Diffusion of Innovations' in 1962 and the author described diffusion to follow an S-shaped curve, time being presented on the x-axis and cumulative per period uptake on the y-axis (illustrated in Figure 1.1).

Figure 1.1 Example of an s-shaped diffusion curve according to Rogers' theory of diffusion



The s-shape resulted from the assumption that populations are heterogeneous in their propensity to innovate (Meade and Islam, 2006), with innovators having a relatively small

threshold to technology adoption and imitators having a relatively higher threshold. Rogers assumed that the threshold sizes were distributed normally among the population (Rogers, 2003) where the size of individual thresholds could depend on various factors, among them income. This model mainly applies to the purchases component of uptake but it has also been assumed to hold for the combined levels of purchases and utilisation by clinicians in the context of health technology diffusion and productivity growth in the United States (Skinner and Staiger, 2015). When implementation appears low in the first periods after technology introduction, this may therefore simply reflect that the technology is early in its product life cycle. This is not to say that nothing can be done to increase implementation. For this, it is important to understand factors that can potentially influence diffusion.

1.2.4 Technology diffusion in the health-care sector

Health technology implementation varies between different technologies and in many countries around the world (Packer et al., 2006), as well as in England (Department of Health, 2011). Low implementation is recognised as a problem in England and is currently on the political agenda (Department of Health, 2011). The Department of Health (DH) committed to a variety of actions to aid the implementation of new technologies such as the introduction of the NICE Compliance Regime that has the aim to ensure rapid consistent implementation throughout the NHS (Department of Health, 2011). There is now a wide range of policy initiatives in use in the NHS to improve implementation, from the provision of financial incentives within the Quality and Outcomes Framework (QOF) and Commissioning for Quality and Innovation (CQUIN) scheme through to the NICE implementation programme that provides local organisations with tools such as 'How to put NICE guidance into practice' (Essat et al., 2013).

To be able to devise effective implementation strategies, it is essential to understand the causes of low implementation. Whilst the Department of Health's plans target the removal of systemic barriers to implementation and incentivise implementation through reward schemes, there are many possible causes for low implementation that may not be tackled through these broad schemes. These include technology characteristics that may lead to low implementation (Terris-Prestholt et al., 2016), patient characteristics such as deprivation (Morris et al., 2012) or financial barriers (Skinner and Staiger, 2015). As diffusion of health technologies appears to be heterogeneous, there is no single set of diffusion determinants that explains low implementation for all technologies, or health technologies for that matter (Greenhalgh et al., 2004). With health technologies being the focus of this thesis, I

concentrated on finding diffusion drivers and barriers from the health technology sector, thereby excluding generic sources applicable to all types of technologies such as the catalogue of technology uptake determinant factors created by Rogers (2003). I performed a scoping review of the literature on uptake and diffusion of health technologies in the United Kingdom (UK), which helped me inform the research questions in this thesis. I identified three exemplary sources with findings on diffusion determinants in health technologies potentially relevant to this work: one was a systematic review of so-called diffusion determinants conducted by Greenhalgh et al. (2004), the other the previously mentioned DH report (Department of Health, 2011), which identified the main barriers to implementation in the NHS and the third was a qualitative study on health technology diffusion determinants in the England NHS (Barnett et al., 2011). These three studies provide insights that are valuable for study design and I therefore briefly summarise their methods and findings:

Greenhalgh et al. (2004) developed a conceptual model for determinants of diffusion in health service delivery and organisation based on their systematic literature review of diffusion determinants in health service delivery innovations. The authors used a broad definition of innovation in health service delivery and organisation as *'novel set of behaviours, routines and ways of working that are directed at improving health outcomes, administrative efficiency, cost-effectiveness or users' experience and that are implemented by planned and coordinated actions'*. Based on the results, the authors examined different theoretical approaches to diffusion of innovations in order to create a conceptual diffusion model. The respective conceptualisation of diffusion of each theoretical approach was extracted. The authors developed a model of diffusion and provided a catalogue of categories of diffusion determinants. The high-level categories included are shown in Table 1.1.

Barnett et al. (2011) conducted a qualitative study in which the authors interviewed fifteen health care organisations, which had generated and implemented a service innovation and were recognised with an award. The aim was to identify diffusion determinants in the England NHS as perceived by the innovators and main themes are shown in Table 1.1.

The Department of Health report 'Innovation Health and Wealth' (2011) summarised the practical experience from the England NHS in terms of barriers to the use of innovations, as shown in Table 1.1.

Table 1.1 The drivers of health technology diffusion according to three studies

Greenhalgh et al. (2004)	**Barnett et al. (2011)**	**Department of Health (2011)**
Characteristics of innovation: its relative advantage over existing interventions as well as its requirements in terms of knowledge and tasks required for adoption	The role of evidence	Poor access to evidence, data and metrics
Adopters: needs, skills, values and goals, and social networks	The role of partnerships	Insufficient recognition and celebration of innovation and innovators
Communication and influence: social networks and the influence of change agents	The influence of champions and other human-based resources	Financial levers do not reward innovators and can act as a disincentive to adoption and diffusion
System antecedents: organisation structure, absorptive capacity for new knowledge and receptive context for change	The impact of contextual factors, both organisational and external	Commissioners lack the tools or capability to drive innovation
System readiness: need, advocacy and opponents and the ability to evaluate the innovation		Leadership culture to support innovation is inconsistent or lacking
Outer context: interorganisational networks, knowledge exchange, international spread strategies and policies		Lack of effective or systematic innovation architecture
Implementation and routinisation: organisation structure, leadership and funding		
Linkage among model components: characteristics of links between change agency, developers and / or potential external change agents		

Owing to the different methods and inclusion of different sources, it is difficult to directly compare these three studies. There are similarities in the findings, but differences are especially evident in the level of comprehensiveness, with the Greenhalgh et al. (2004) study being far more comprehensive than the other two; and the level of abstractness, with the DH report being the least abstract of the three. Furthermore, the DH report focussed on effective

and even cost-effective technologies experiencing low implementation, and therefore ignored some of the factors mentioned by Greenhalgh et al. (2004), such as the technological advantage.

This brief review demonstrated that these exemplary 'generic' catalogues cannot readily be used for identifying diffusion determinants individual to specific technologies as diffusion is a highly technology-specific process (Greenhalgh et al., 2004). The above catalogues present an aggregation of individual factors affecting uptake of a number of individual health technologies and, as such, present inspiration for what may affect uptake of an individual technology. However, no final conclusive decision on relevant uptake factors to an individual technology can be reached using these catalogues. Apart from this individuality, the development stage that the technology in question is at may affect what is regarded as most important uptake factors and has to be borne in mind when devising implementation strategies. For instance, a technology's research evidence on effectiveness is likely to influence uptake at any stage of development whereas usability may only present a barrier to uptake once research evidence is available to show that the technology works. These catalogues also do not tell us anything about the design of potential implementation strategies.

1.2.5 Health technology assessment methods and diffusion

It appears as though implementation estimates are not routine components of analyses performed within health technology assessments (NICE, 2013), with the potential exception of the requirement for a budget impact analysis. Budget impact analyses identify an estimate of the total expenditure associated with a health technology. The population estimate used to arrive at such an estimate should be adjusted by the current technology mix in use, as was highlighted in the ISPOR guide to budget impact analysis (Mauskopf et al., 2007). Mauskopf et al. (2007) recommended obtaining data to inform this from the decision-maker's own database and to account for changes over time. They did not specify any method as to how the changes over time can be accounted for. They further recommended use of market research data or expert opinion on current and evolving treatment patterns in the absence of published information on current treatment patterns (Mauskopf et al., 2007).

Another area in which implementation estimates have appeared in health economic analysis is the assessment of implementation strategies. Economic evaluations of implementation strategies are relatively uncommon (Essat et al., 2013) and are not part of standard health technology assessment processes (NICE, 2013, NICE, 2012, NICE, 2011a).

Essat et al. (2013) identified frameworks for the evaluation of implementation strategies (Fenwick et al., 2008, Mason et al., 2001, Sculpher, 2000), and identified only a few further developments and applications (Willan and Eckermann, 2010, Walker et al., 2010, Hoomans et al., 2009a).

NICE methods guides do not recommend the use of implementation estimates within cost-effectiveness analyses (NICE, 2013, NICE, 2012, NICE, 2011a). Most health economic models focus on modelling one cohort over the lifetime of patients (Hoyle and Anderson, 2010). The implicit assumption is that no change in parameter values occurs over the modelled decision relevance horizon, thereby ignoring potential correlations between changing implementation rates and other model parameters.

1.3 Rationale for this thesis

1.3.1 Why does diffusion matter in health technology assessment?

In the background section, I highlighted that implementation of effective health technologies is often low. There are generic causes for low implementation although these vary for individual technologies. Health economic methods, at least in England, do not routinely incorporate estimates of diffusion. For the health system, or the health-care payer, however, low implementation of cost-effective technologies results in an inefficient use of resources. This is because cost-ineffective technologies remain in use when cost-effective technologies are not fully implemented, causing these inefficiencies. Furthermore, health economic models typically ignore future change. The question that came to my mind was: what are the implications of health technology diffusion for how we value health technologies?

Four hypotheses motivated this thesis: I hypothesised (A) that consideration of low implementation and the resulting opportunity loss to the health care system could be beneficial to the reimbursement authority and the health care payer at the time of HTA, by enabling decision-making on implementation strategies, especially the generation of further research evidence; (B) that implementation estimates may be correlated with other parameters in the cost-effectiveness model and influence health economic outcomes; (C) that it is possible to acquire the knowledge of low implementation and the opportunity loss it causes at the time of HTA; and (D) that a better understanding of the technology-specific causes of low implementation is required to effect change.

Hypothesis (A) is derived from the thought that consideration of the opportunity loss caused by low implementation at the time of HTA may enable the payer or the decision-maker to address the problem of low implementation early on, that is, before the technology is introduced into routine practice. The decision-making authority may be able to recommend implementation strategies, although NICE, for instance, can only make recommendations for further research to be conducted and not for other implementation strategies to be pursued (NICE, 2013). It therefore makes sense to examine research recommendations for their ability to have a positive effect on implementation.

Hypothesis (B) refers to the potential implications for HTA where the implementation parameters are correlated with other parameters in the health economic model. If this is the case and the implementation parameter changes over time, there may be a case for incorporating these changes and their impact in the model. One such case is the effect of diffusion on price. It has been observed in medical devices that prices declined with increasing technology uptake, a phenomenon that is described by so-called experience curves (Brown et al., 2007).

To study hypothesis (C), it is important to understand the opportunity loss to the payer that is caused by low implementation. This has previously been called the expected value of perfect implementation by Fenwick et al. (2008) and is a function of the implementation discrepancy (that is, how low implementation is compared to the desired 'full' level) and the expected incremental net benefit of the cost-effective technology over its alternatives. Whilst the former, the estimation of the implementation discrepancy, is not a routine part of HTAs, the estimation of the expected incremental net benefit is. Estimating the value of implementation at the time of HTA would provide the decision-maker with valuable information early on, with the only additional requirement of estimating the implementation discrepancy.

Hypothesis (D) is related to developing an understanding of the technology-specific causes of low implementation that is required to effect change in implementation. Knowledge of the causes of low implementation can potentially enable stakeholders to design effective implementation initiatives. Such knowledge is relevant for technology developers and policy-makers alike. The questions that may be asked by policy-makers and technology developers are: what are barriers to uptake, how can they be addressed and what is the cost-effectiveness of addressing them? Answering these questions poses a challenge: when a technology is still in the development phase, uptake barriers have to be identified prospectively before any uptake could be observed. Furthermore, there may be an abundance of factors that play into the uptake decision of a diverse set of different individual and

institutional decision-makers. To identify technology-specific diffusion determinants and implementation strategies in a certain environment and before the HTA process, technology-specific research prior to technology launch is required. As was described above, there do not appear to be any prescribed methods to find out about technology-specific uptake factors prospectively.

1.3.2 Application in a case study

The hypotheses developed in the previous section were tested in this thesis using a real life case study of a technology currently in development, called electrical impedance spectroscopy developed for preterm birth screening. The use of only one case study may be perceived as a limitation but I accepted this to ensure feasibility of the project within the time frame of a doctoral thesis while addressing all of the above-mentioned hypotheses. The choice of the case study technology was largely dictated by access to data of technologies in development. This proved easier for technologies other than drugs. I thought it adequate to choose a medical device over a drug because the DH report (Department of Health, 2011) predominantly mentioned 'non-drugs' for which particularly low implementation was noted.

1.3.2.1 Therapeutic area of preterm birth

In the UK, approximately 4% of all babies are born prematurely, that is before 34 weeks of gestation (Honest et al., 2009). Preterm birth (PTB) is a serious issue associated with potential health consequences for the baby, encompassing neo-natal mortality, long-term neurological impairment, respiratory distress syndrome, bronchopulmonary dysplasia, intraventricular haemorrhage, retrolental fibroplasia and developmental problems (Honest et al., 2009). PTB places a great emotional, psychological and financial burden on parents and their children and is associated with significant cost, chiefly for prolonged neo-natal care and health care.

There is a variety of causes of PTB, some of which are not fully understood. Some cases of PTB are elective, which means that it is judged to be in the baby's and mother's interest to deliver the baby early due to complications such as hypertension or diabetes (Honest et al., 2009). The non-elective cases are the focus of this thesis but even these are heterogeneous and the cause of a PTB in an individual is not always identified. Causes include inflammations, cervical weakness, and other environmental factors, amongst others, with cervical weakness being in the focus of this thesis.

Owing to the heterogeneity of the condition, there are many tests (a total of 22 as per the review by Honest et al. (2009)) that claim to predict PTB, with varying levels of accuracy. These include a previous history of PTB, cervicovaginal fibronectin, cervical length ultrasound scanning, amongst others. Not all of these specifically test for cervical weakness. If a pregnant woman received a positive test, then there is a range of 40 different treatment / prevention options (Honest et al., 2009). Many of these treatment options lack a good quality evidence base, but some were associated with a benefit in preventing PTB, including smoking cessation programmes, progesterone treatment, periodontal treatment, fish oil (Honest et al., 2009) and cervical cerclage (NICE, 2015). The current NICE recommendation for women with potential cervical weakness is to offer a choice of either prophylactic vaginal progesterone or prophylactic cervical cerclage to women with a history of PTB or mid-trimester loss and in whom ultrasound scan has revealed a cervical length shorter than 25 mm between weeks 16 and 24 of pregnancy (NICE, 2015).

There are, however, limitations associated with cervical length ultrasound scans. One is that the predictive ability is not very good: its sensitivity was estimated at approximately 21% and its specificity at 98% (based on likelihood ratios provided by Honest et al. (2009)). This means that cervical length ultrasound scans are not very good at identifying those women that will go on to have a PTB, as only 21% of those that will have a PTB test positive. A Cochrane review (Berghella et al., 2009) has therefore highlighted the need for new and improved screening methods. Another limitation is that ultrasound machines are often working at full capacity, even without them being used for PTB screening (NICE, 2015).

1.3.2.2 Electrical Impedance Spectroscopy for use in PTB screening

Electrical Impedance Spectroscopy (EIS) is a technology that has been developed and used for cervical cancer screening. In this function, it is manufactured by Zilico Ltd. It is also currently in development for use in PTB screening at Sheffield Teaching Hospital. EIS measures cervical impedance, which has been observed in early tests to be predictive of PTB before 34 and 37 weeks of gestation in women at risk based on their previous history. EIS examinations are thought to occur at week 20-22 and week 26-28 of a woman's pregnancy. A randomised controlled trial in women with a history of PTB is currently under way to establish the predictive ability of EIS. Preliminary results from a pilot based on a sample size of 38 women only indicate that EIS has a sensitivity of 44% and specificity of 90%, indicating that there may be a slight improvement over cervical length ultrasound scans in the reduction of false negatives. It would provide the additional advantage of being an easy to use, small and probably inexpensive piece of kit that could alleviate the pressure on

ultrasound machines. The developers envision that the main treatment strategy following a positive test result would be treatment with vaginal progesterone. One issue for the development of EIS for the purpose of PTB screening was that the device's accuracy was sensitive to pressure exerted by the EIS administrator, which leads to measurement variability. Therefore, the device is being further developed by adding a transducer-based force gauge that may improve test accuracy over what was observed in the pilot study.

1.3.3 Research philosophy

A researcher's philosophical position influences their methodology and is thus viewed as important to identify before embarking on a research project (Ritchie et al., 2003). My intuition was to use different methods in this thesis and I therefore thought it valuable to identify my own research position and how it influences the choice of methodology. A philosophical position is typically characterised by the stance a researcher adopts with regards to the paradigms epistemology and ontology (Ritchie et al., 2003). For this study, I adopted an epistemological stance of pragmatism with the ontological view of subtle realism.

Ontology is the study of the nature of all being. It comprises a range of different positions between two extremes: realism acknowledges the existence of one reality that is independent of anyone's beliefs, while idealism states that reality is only knowable through socially constructed meanings (Mason, 2002). A position between the two is subtle realism, which acknowledges the existence of a reality independent of people that is only accessible through people (Ritchie et al., 2003). In the context of this thesis, the stance of subtle realism signifies that while I believe in the existence of one reality and I am striving to capture this reality in a health economic modelling framework and through other research methods, I acknowledge that my understanding of it may be shaped by assumptions that are a product of my priorities and beliefs.

Epistemology refers to the question of how knowledge is created. There are the two extremes of positivism and interpretivism, where positivism describes the researcher's stance that created knowledge is independent of and unaffected by the researcher (Mason, 2002), resulting in a preference for quantitative research methods. With the stance of interpretivism, in contrast, the created knowledge of the social world only exists through the interaction with the researcher, leading to qualitative research being viewed as appropriate (Mason, 2002). The choice of pragmatism as a paradigm expresses the rejection of a forced choice between the two and enables research methods to be chosen based on the research problem, rather than being dictated by the chosen position itself (Ritchie et al., 2003). The

rationale for choosing the pragmatist stance was that I believe that different research questions warrant different methods.

1.4 Research aim and objectives

The aim of this thesis was to develop a method to estimate health technology diffusion prior to the time of HTA and to explore the potential impact of these diffusion estimates on HTA methods. To achieve this, I sought to address the following objectives:

1. To explore the potential use and impact of diffusion estimates in the health technology assessment context.
2. To identify prediction methods for diffusion prior to technology introduction and relevant to the health technology context that yield probabilistic diffusion estimates.
3. To develop model frameworks for incorporating diffusion estimates in relevant health economic analyses.
4. To identify relevant implementation strategies for the case study technology EIS for use in PTB screening.
5. To develop a method for predicting diffusion prior to technology introduction using the PTB screening case study.
6. To assess the impact of diffusion on health economic model outcomes in the PTB screening case study.

1.5 Thesis structure

The thesis structure is described below and illustrated in Figure 1.2. Chapter 2 addresses Objective 1 through a series of literature reviews. These reviews comprise a scoping review to identify common uses of implementation estimates in health economic evaluation; a second review on the application of experience curves in health economic modelling; a third review on the use of implementation estimates in value of information analyses; and a fourth

review on the status quo of applications and developments of the expected value of implementation.

In Chapter 3, two literature reviews are presented that address Objective 2. The first review aims to collect diffusion data from the UK health technology setting in order to assess the feasibility of evidence synthesis. The second review aims to identify a method for predicting diffusion prior to technology introduction.

In Chapter 4, I present model frameworks for the potential effects of diffusion estimates on health economic analyses to address Objective 3. The first framework develops the value of implementation further by allowing for integration of the dynamics of diffusion and by allowing information to affect implementation. The second framework establishes the relationship between health technology implementation and price.

Chapter 5 describes the qualitative study that was conducted to identify potentially relevant implementation strategies for the case study technology (Objective 4). Within this, I report on semi-structured interviews that were conducted with participants involved with the use or purchase of health technologies in obstetrics to identify diffusion determinants. Relevant diffusion determinants were then translated into two implementation strategies.

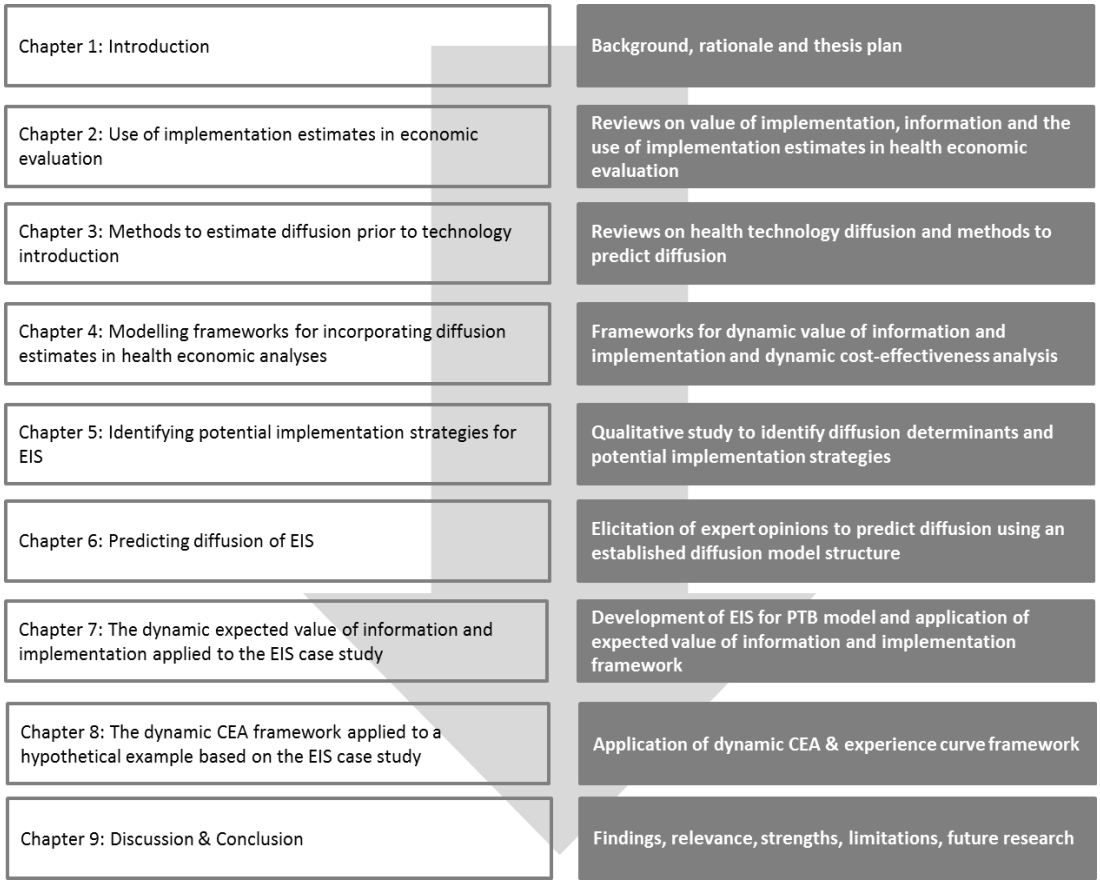
In Chapter 6, I describe the developed diffusion prediction method and its application to the case study technology (Objective 5). The developed method was an elicitation of expert opinion using the structure of an established diffusion model.

In Chapter 7, the expected value of information and implementation framework that is described and extended in Chapter 4 is applied to the health economic analysis of the case study technology (Objective 6). First, the cost-effectiveness model of EIS is presented. The values of implementation, information and further research are then calculated.

Chapter 8 presents the application of the dynamic cost-effectiveness analysis and experience curve framework to a hypothetical case study based on the EIS case study setting (Objective 6), as it turned out that the cost-effectiveness results of EIS were insensitive to price.

In Chapter 9, I discuss the novel contribution of this thesis in the context of the existing literature. I present how this thesis is relevant and discuss strengths, limitations and areas for further research. I finally provide a conclusion.

Figure 1.2 Thesis flow



CHAPTER 2. REVIEWS ON THE POTENTIAL USE OF IMPLEMENTATION ESTIMATES IN HEALTH ECONOMIC EVALUATION

2.1 Background

Chapter 1 highlighted that little was known about the relevance of implementation estimates to health technology assessment and methods for estimating implementation for use in health economic models. The aim of this chapter is to explore the uses of implementation estimates in health economic evaluation. Four reviews were performed to address this. One aimed to identify common uses of implementation estimates in economic evaluations of health technologies. The other three reviews more specifically targeted defined uses of implementation: identifying health economic evaluations that had modelled experience curves, exploring the use of implementation estimates in *Expected Value of Information (EVI)* analysis and investigating the use of the *Expected Value of Implementation (EVIM)*.

2.2 Implementation estimates in health economic evaluations – findings from an abandoned scoping review

2.2.1 Background to implementation in economic evaluation review

This literature review was conducted to identify common uses of implementation estimates in health economic evaluations, to find the reasons for including implementation, the methods used and to examine what effect implementation had on model results. This scoping review was started at the very beginning of this project and it quickly turned out that it was not suited to define the research questions for this thesis. This was mainly because a search on methods is typically difficult (more on this in Section 2.2.2). I therefore abandoned the review half-way through, to concentrate on more specific research questions, which are described and answered in Sections 2.3 to 2.5 of this chapter. However, this

abandoned review, while not comprehensive, resulted in some interesting findings. I therefore decided to include the review in this thesis.

2.2.2 Methods of implementation in economic evaluation review

A scoping review as described by Grant and Booth (2009) was considered appropriate to find out more about implementation estimates used in health economic methods. This review type identifies the nature and extent of research evidence on a subject, informing the researcher what additional reviews may be needed. In scoping reviews, there is some flexibility as to the level of completeness of the review. In the context of this thesis, the scoping review could help identify areas for further research. These potential topics could then be researched more efficiently with more targeted reviews. Methodology papers are notoriously challenging to identify: searches typically yield a large number of hits, of which only a few may be relevant (Chilcott et al., 2003). Citation searches or searching lists of references of key publications are a more appropriate approach to search for further documents on a subject where there are known deficiencies in indexing or terminology (Chilcott et al., 2003).

2.2.2.1 Search strategy

The search was conducted in Medline via OvidSP. The search terms were selected such that only studies on economic evaluations in health were considered that had concerned themselves in any way with implementation. To capture potential health economic evaluations, the terms HTA (for health technology assessment) and CTA (for constructive technology assessment, a concept developed in the Netherlands where implementation estimates play a role in some economic evaluations known to the author) were used alongside other terms that are commonly associated with health technology assessments. Largely synonymous terms with implementation were also used in the search strategy. The following building blocks were used:

Facet A: (Implementation OR uptake OR diffusion)

Facet B: (Model OR estimate)

Facet C: (HTA OR CTA OR cost effectiveness OR economic evaluation OR technology assessment)

Search: (A AND B AND C)

In the first phase, titles and abstracts were screened; full texts of the included studies were screened in the second phase.

2.2.2.2 Inclusion and exclusion criteria

Inclusion criteria were economic evaluations of health technologies that included an estimate of implementation. Studies that evaluated the effectiveness of initiatives to increase uptake were excluded, as these were dealt with in a separate review.

2.2.2.3 Data extraction

A combination of tabular and narrative analysis was used. The standardised tabular data extraction sought to provide an overview of the study characteristics that could potentially address the research questions. Answers to the research questions provided by this review were summarised in the narrative:

1.) What were the technologies under assessment in the identified studies?

2.) What was the impact of the implementation estimates on health economic outcomes?

3.) What were the methods used to estimate implementation?

4.) What areas for further research on implementation were identified?

2.2.3 Results of implementation in economic evaluation review

This search, which was conducted in May 2012, identified 515 publications. Two hundred of these publications were reviewed. The 200 reviewed publications were a pseudo-random sample from the 515 hits, using the first letter of first authors' surnames. At least 10% of hits for each first letter were reviewed. Of these, 163 publications were excluded based on titles and abstracts. Of the remaining 37 hits, 17 were excluded based on a full text review, leading to 20 included articles. One further publication, not identified through the search, was also included: a follow up publication on one of the studies found in the review (Entry 19) (Retel et al., 2012). An overview of the included publications is shown in Appendix Table A.1. As mentioned above, this review was abandoned after 200 hits. At this point,

only few uses of implementation estimates in cost-effectiveness analysis citations were found, thus highlighting the difficulty of finding relevant publications with this type of search. Whilst the remaining 315 would have included further relevant citations, I thought that I could answer my research questions better in the reviews described in Sections 2.3, 2.4 and 2.5. This is a consequence from the results, which are described in the next Section 2.2.3.

The results of this scoping review suggested that estimates of implementation were rarely used in economic evaluations. Out of the 200 hits reviewed, only 10 percent were economic evaluations that had included an implementation estimate. However, as per the definition of the search, only economic evaluations that had mentioned implementation were included in the review. The proportion of economic evaluations with implementation estimates is therefore unclear, but is likely to be much lower than 10 percent.

1.) What were the technologies under assessment in the identified studies?

Out of 20 identified studies reported in 21 citations (as stated above, a follow-up publication (Retel et al., 2012) on the same study by Retel et al. (2009) was also included), 14 were economic evaluations of infectious diseases, of which 12 were evaluations of vaccinations. Herd immunity was included as an outcome in three of these vaccination studies. Seven of the identified publications evaluated screening or diagnostic technologies, three of which were in infectious diseases and the remaining four in other conditions. In addition, one study evaluated a disease-management programme in coronary heart disease, and one study assessed different chemotherapy regimens.

2.) What was the impact of the implementation estimates on health economic outcomes?

In 12 studies reported in 13 publications, the implementation estimate had an effect on the *Incremental Cost-Effectiveness Ratio (ICER)* (Entries 1, 2, 3, 6, 7, 8, 10, 11, 15, 16, 19, 20, 21 in Appendix A.1). Three studies included implementation estimates only to obtain estimates on population effects and / or costs (Entries 5, 17, 18 in Appendix A.1). In three studies, the impact of implementation was not clear (Entries 4, 13, 14 in Appendix A.1). One study evaluated the expected value of implementation (Entry 9 in Appendix A.1). Another study estimated the cost per affected pregnancy (Entry 12 in Appendix A.1).

The mechanism by which implementation estimates affected the ICER in the 13 identified publications varied. In three of the 13 publications in which implementation had

an effect on the ICER, the mechanism was by which this occurred was herd immunity (Entries 7, 10, 17 in Appendix A.1). With herd immunity, effectiveness grows at a greater rate with implementation than costs. A fourth study (Entry 21 in Appendix A.1), reported a similar mechanism by which the ICER varied with implementation, by effects growing at a greater rate with implementation than costs. While herd immunity was not specifically mentioned, the effects of the different chlamydia screening programmes were considered beyond the screened individual in a transmission model (Roberts et al., 2007a).

In three further studies reported in four publications, the levels of implementation affected the ICER because implementation levels differed for the different comparator strategies and the ICER was calculated as a so-called 'blended ICER' in which costs and effects of each strategy were adjusted by the implementation levels of that strategy (Entries 1, 3, 19, 20 in Appendix A.1). For instance, in the cost-effectiveness analysis of influenza vaccination in 50-64 year olds by Aballea et al. (2007) (Entry 1), uptake would generally be higher among high risk patients than among low risk patients. A cohort vaccination strategy would result in even higher uptake among the high risk population that exhibited a higher propensity to benefit from vaccination compared to a high risk only vaccination strategy. Because each strategy was associated with different implementation levels, the ICER was affected by the implementation estimate. Another study tested different uptake scenarios for baseline care, in this case a community eye test, to estimate the cost-effectiveness of open angle glaucoma (OAG) screening (Burr et al., 2007) (Entry 3 in Appendix A.1). When higher uptake rates of the community eye test were employed, the ICERs for OAG screening compared with the community eye test were larger because the cheaper community eye test would result in many OAG cases detected without screening for OAG specifically. One study that was reported in two included articles on the same constructive technology assessment (CTA) study on a breast cancer diagnostic test in the Netherlands focused on forecasting future developments, particularly diffusion (Retel et al., 2009, Retel et al., 2012) (Entries 19, 20 in Appendix A.1). This is a requirement of CTA studies and the rationale for this is to provide information to potentially influence future developments. The blended ICER for the new technology was therefore higher when uptake was lower, due to the less cost-effective comparator remaining in use when uptake was low. In another study by Atherly et al. (2009), on the cost-effectiveness of rotavirus vaccination compared to no vaccination, the mechanism by which the ICER varied with implementation was the patient mix of high risk and low risk groups with different propensity to benefit and different uptake rates (Entry 2 in Appendix A.1).

Among the four remaining publications out of the 13, there were three for which the ICER was said to be affected but the mechanism by which this occurred was unclear (Entries 8, 15, 16 in Appendix A.1). Authors were contacted but no reply was received.

The remaining publication was a study on a management programme for coronary heart disease (Chew et al., 2010) (Entry 6 in Appendix A.1). The authors found uptake relevant because full costs of the programme would be incurred for all patients, including those who did not complete the programme; while effects would not be accrued by those non-compliant patients. With varying uptake, the rate of non-completions would also be varied, presumably not proportional to the change in uptake, although this level of detail was not provided.

3.) What were the methods used to estimate implementation?

The methods of obtaining implementation estimates were relatively basic. Thirteen studies used historical data or data on analogous technologies, for example, general infant vaccination coverage rates. These assumed rates did not change over time. Out of those 13, four studies made adjustments to the available data based on assumptions. Six studies based their implementation estimates on assumptions only. One study used diffusion theory in discussions with experts, created qualitative diffusion scenarios with the help of these experts and based the quantitative estimates on various assumptions (Retel and Grimm, 2013). For example, in one scenario, uptake would resemble that of uptake of clinical guidelines, in another scenario it would be at 50% for the entire duration of decision relevance.

4.) Was any area for further research on implementation identified?

Twelve out of the 21 publications did not mention implementation in their discussion. One study highlighted that the implementation estimate was only used for the *Expected Value of Perfect Implementation (EVPIM)* but not for the ICER or EVPI calculations for which it may be beneficial (Entry 9 in Appendix A.1). Two studies highlighted the potential for cost-effectiveness evaluations to assess uptake-increasing initiatives (Entries 1, 21 in Appendix A.1). Five studies acknowledged limitations with the way implementation levels were estimated and added that future research should focus on obtaining better quantitative estimates (Entries 3, 6, 11, 16, 20 in Appendix A.1). One study acknowledged the static nature of their implementation estimate of pertussis vaccination as a limitation (Entry 10 in

Appendix A.1) and another discussed the potential for incorporating differential uptake rates for different patient characteristics in the chlamydia screening study (Entry 21 in Appendix A.1).

2.2.4 Discussion of implementation in economic evaluation review

This scoping review found that approximately 10 percent of the search results were relevant to address the broadly defined research question, suggesting that, if completed, this review could have resulted in approximately 50 relevant publications. The review highlighted that it was likely that implementation estimates were not commonly used in health economic evaluations. Implementation estimates were used in those types of economic evaluations in which implementation affected health economic outcomes, such as in infectious diseases through herd immunity for instance, or value of implementation analyses. Importantly, the methods used to estimate implementation were simple: either data from reference technologies were used, or the estimates were based on assumption.

These preliminary results suggested that there was not much added value in finishing the review or conducting a systematic review on the topic. This was because results did not do a particularly good job in defining further research questions. The findings from this review include that the inclusion of implementation estimates appeared to be fairly established in infectious disease modelling and in the estimation of population costs and effects, suggesting that there was limited scope for further research in this area, except in methods to obtain better implementation estimates. The ICER was predominantly affected when a blended ICER was calculated, where costs and effects of each comparator were adjusted by levels of implementation that differed for the comparator. Such a blended ICER makes intuitive sense in technologies for which effectiveness relies on the level of implementation. However, blended ICERs do not identify the most efficient technology based on their costs and health effects, which is arguably the prime objective of reimbursement authorities such as NICE. Blended ICERs are therefore not recommended by NICE (NICE, 2013).

Another topic that was excluded as an area for further research was the inclusion of patient groups with different implementation levels. Subgroups or patients with different characteristics are better dealt with in separate economic evaluations (Sculpher, 2008). If this were not done, a new technology may be considered cost-effective on average, despite it being cost-ineffective in some subgroups (Sculpher, 2008). The distinction between patient compliance to treatment and implementation was another relevant issue identified. This thesis focusses on implementation, not compliance, and, therefore, non-compliance to

treatment programmes was not included as an area for further research. The expected value of implementation was identified as one area where further research may be beneficial and this is pursued in Section 2.5. No value of information analysis studies were identified in this review.

The limitations of this scoping search were that it was confined to one database, Medline, and that it was abandoned after 200 search results were sifted through. I have provided justification for this in that there is not expected to be added value in continuing this review and that other review designs would be better in identifying relevant publications. However, the lack of comprehensiveness may mean that potentially interesting areas for further research were not identified.

In conclusion, conventional review methods provide little value for methodological questions such as this one and pearl-growing and citation searches are generally considered more useful (Chilcott et al., 2003). However, this review has helped in identifying that methods of estimating implementation for use in health economic modelling were likely to be rudimentary and that there were potential areas for further research in the use of expected value of implementation analysis.

2.3 Review on the use of experience curves in health economic evaluation

2.3.1 Background to the experience curve review

The previous review showed that implementation levels are not commonly used in economic evaluations. One reason for this might be that the impact of implementation on commonly used measures such as the ICER is not immediately obvious. For the standard (that is, the non-blended) ICER to be affected by implementation levels, implementation would have to affect costs and effects in a disproportionate manner. The above review showed that this could be the case when there is herd immunity to be achieved in infectious diseases. Another way this may occur is through the existence of price-volume relationships, which are described by so-called experience curves that were introduced in Chapter 1.

Experience curves describe empirical evidence on price declines with increasing uptake of technologies. Experience curves were first described in a study by Wright (1935) but since then, ample evidence for such price reductions in a variety of different technologies as

well as from a study on 20 medical devices by Brown et al. (2007) have become available. The existence of experience curves is mainly explained through a technology's competitive situation (Brown et al., 2007). When the conditions of perfect competition and perfect information are not satisfied, pricing occurs above marginal costs, especially in R&D intensive industries (Camejo et al., 2011). The larger a market becomes, the more likely it is for competitors to enter. In the health-care industry, this would typically occur after patent expiry but also before, through between-patent competition, which describes competition through close substitutes (Camejo et al., 2011). With increasing competition, prices are likely to fall. In addition, economies of scale, which describe reductions in costs with increasing production volume at the level of the market actor may also lead to reduced costs and prices (Brown et al., 2008, Brown et al., 2007). As both, the competitive situation and economies of scale (to a lesser extent as most of the health technology price would typically reflect development costs) may hold for drugs, experience curves may be applicable to drugs, although there is no evidence of such to my knowledge.

The findings presented by Brown et al. (2007) imply that there may be value in including a dynamic model of product price in health technology assessments of medical devices. If this is not done, a technology may be rejected based on its initial price while, after a certain level of diffusion, the product's final price may be lower. Importantly, these price changes would only materialise with a positive recommendation – if the technology is not implemented, no price changes will occur. If experience curves were present, implementation would thus influence price and cost-effectiveness. The question whether this relationship has been considered in economic evaluation is addressed with this review.

2.3.2 Methods of the experience curve review

I performed a citation search to identify those articles that cited the study by Brown et al. (2007) on medical device prices following the experience curve. The rationale behind this was that any study after the publishing date of Brown et al. (2007) that had used experience curves in health economic modelling would presumably have cited this study.

2.3.2.1 Search strategy

I performed a search in three search engines: 1. Medline, 2. Web of Science and 3. Google Scholar. I performed citation searches of the key journal article by Brown et al. (2007) in all three search engines.

2.3.2.2 Exclusion criteria

Excluded were all search results that were (A) not a health economic model, (B) that had not used experience curves within the model or (C) were not in English.

2.3.2.3 Data extraction and analysis

Search results were screened based on their titles and abstracts in a first phase and based on their full texts in the second phase. Included publications were described in a narrative.

2.3.3 Results of the experience curve review

The search that was conducted in November 2015 produced 12 results. All of these were obtained from the Google Scholar search as the identified key publication by Brown et al. (2007) was not found in Medline or in Web of Science. Eight of the 12 results were excluded based on exclusion criterion (A), they were not health economic models, in the first title and abstract screening phase. One study was excluded based on exclusion criterion (C), with title, abstract and full text only available in Japanese.

These led to three publications being screened in the second phase full text review. All three were excluded, one because it was not a health economic model (A) and the other two because they did not use experience curves within the model. One of these mentioned potential price declines in the discussion (Osnabrugge et al., 2012). The other publication mentioned potential future device price fluctuations as a reason for performing deterministic sensitivity analysis on device price (Moreno et al., 2012). Experience curves were, however, not used in the model.

2.3.4 Discussion of the experience curve review

This brief review showed that potential future price changes, as supported by evidence on experience curves in medical devices, have so far not been included in health economic evaluations. There may therefore be value in exploring methods of how they can be incorporated in economic evaluation of health technologies and exploring potential cost-effectiveness results.

The limitation of this review was its small scope, using a publication as a starting point that was not widely cited. However, it is not very likely that an economic evaluation in health technologies posterior to the Brown et al. (2007) article would not have cited this study, as this appears to be the only study on experience curve data in the health technology sector.

In conclusion, the inclusion of experience curves in health economic evaluation has not been widely studied and presents a valuable area for further research within this thesis.

2.4 Review of implementation estimates in value of information analyses

2.4.1 Background to implementation in EVI analysis review

The review in Section 2.2 did not result in identification of any value of information analyses. However, there is the potential for implementation to affect the value of information, particularly through the population estimates that are commonly used to compare the population value of information with the cost of research. While the use of implementation-adjusted population estimates for costs and effects within cost-effectiveness was documented in the review described in Section 2.2, there was no mention of population estimates for EVI analysis.

EVI is used in economic evaluations to assess the value of reducing decision uncertainty through further research (Briggs et al., 2006). To obtain the overall value that can be derived from research in a health system or society, EVI is aggregated over the population that may potentially benefit and the time horizon that research may be relevant for (subsequently called the decision relevance horizon), resulting in the population EVI (PEVI) (Briggs et al., 2006). There has been some discussion on both, what decision relevance horizon to apply in EVI analyses (Philips et al., 2008) and on methods for obtaining population estimates that should include an estimation of the probability of implementation (Hoomans et al., 2012). With this background, the objective of this research was to investigate whether implementation estimates were used in PEVI analyses.

2.4.2 Methods of implementation in EVI analysis review

2.4.2.1 Search strategy

The search was conducted using the NHS Economic Evaluation Database (EED). Search terms were: ((Value of Information) OR EVI OR VOI). Titles and abstracts were screened in the first phase; and in the second phase, full texts of the thus far included publications were reviewed. Additional PEVI analyses that were known to me but did not come up through this small-scale search were also included.

2.4.2.2 Inclusion and exclusion criteria

Publications were excluded when they did not report PEVI analyses. Duplicates were also removed.

2.4.2.3 Data extraction

A tabular overview of the included studies was developed and results were discussed in a narrative, addressing the following research questions:

1.) Was the population estimate for the PEVI analysis adjusted by uptake?

2.) What methods were used to obtain the estimate of uptake?

3.) Was the issue of uncertainty associated with the population estimate discussed?

2.4.3 Results of implementation in EVI analysis review

The NHS EED search, which was first performed in June 2012 and updated in November 2013, resulted in 43 hits. The titles and abstracts of all hits were screened and 10 publications were excluded because no EVI analysis had been performed and another six were excluded because no population EVI had been reported. Another publication was excluded because it resulted from an analysis already described in another included publication. One publication stated that EVI analysis had been conducted but did not detail methods or results and instead referenced another publication. Hence, the former was excluded and the latter included. Three additional publications with EVI analyses that had

not come up through this search but were known to the author were included. A total of 29 publications were included and examined with regards to the research questions. Results are shown in detail in Appendix A.2 and are described below.

1.) Was the population estimate for the PEVI analysis adjusted by uptake?

In 27 out of the included 29 publications, the population size was not adjusted by uptake of the cost-effective intervention. In those analyses, the population estimate was mainly obtained by estimating the incidence and prevalence of the condition or the number of annual procedures. An uptake level of 100% was implicitly assumed in those cases. Only two studies adjusted the population estimate by uptake.

2.) What methods were used to obtain the estimate of uptake?

One of the two studies mentioned above, on thromboembolism prophylaxis in post-hip replacement patients, used uptake estimates that were based on assumption (McCullagh et al., 2012). The second study used an uptake estimate which was obtained from a trial that had reported the use of trastuzumab adjuvant to chemotherapy in early stage breast cancer (Hall et al., 2011). Neither of the two studies used a dynamic estimate of uptake or accounted for uncertainty associated with it.

3.) Was the issue of uncertainty associated with the population estimate discussed?

Five studies mentioned the issue of uncertainty associated with the population estimates in their PEVI analysis; four in their discussion and one implicitly by reporting different PEVI results with varying population estimates. In one of the five studies, it was stated that implementation may not automatically happen once research results become available and further exploration of this topic was recommended. None of the five studies, however, fully accounted for uncertainty by modelling the population estimate probabilistically.

2.4.4 Discussion of implementation in EVI analysis review

This review showed that the majority of EVI analyses do not consider uptake adjustments in their population estimates. The implication is that most reported values for the PEVI are likely to be an over-estimate of the actual value of further research as low uptake would cause the population benefitting from further research to be smaller than the potentially eligible population. As the PEVI is viewed as an upper ceiling to the value of further research, assuming an uptake level of 100 percent is not wrong per se. It simply reflects the value of further research in a full uptake scenario. The full population EVPI is therefore an important measure. In the context of the value of future research studies, however, this maximum value of further research might never be reached due to low implementation. In cases where an uptake level of 100 percent is unattainable, potentially due to the existence of implementation barriers, ignoring uptake from the PEVI estimate would result in a drastic over-statement of the value of further research and therefore have the potential to mislead decision-makers.

Limitations of this study include the small scope of the search. It is improbable that all PEVI analyses in health technologies have been captured with the adopted search strategy and there may be further examples that incorporated uptake estimates in their analysis, especially after the search was last updated. With 27 out of 29 analyses estimating the population without an uptake adjustment and none including uptake dynamics, the main findings of this review are fairly representative of common practice and could be used as the basis for defining research questions for this thesis.

As was suggested in the previous reviews, this review supports the finding that further research is needed on estimation methods for uptake, which should consider the dynamic nature of implementation and uncertainty. The two studies in this review that had used uptake estimates had obtained them from available trials and assumed that this level of uptake would hold up to the decision relevance horizon. As technology implementation is regarded as a dynamic process, there is further research potential on approaches to estimate uptake dynamics.

In conclusion, based on the result that very few PEVI studies adjusted their population estimate by uptake and taking into account the large downward effect that uptake adjustments could have on the value of PEVI estimates, there is a need for discussion and further research around uptake adjustments in PEVI analyses.

2.5 Review of applications and further development of the expected value of implementation

2.5.1 Background to value of implementation review

The aim of the value of implementation review was to obtain an overview of scope for further development of expected value of implementation methods. Having provided a unified framework of both value of information and value of implementation that is consistent with the methods of technology appraisals, the article by Fenwick et al. (2008) was used as the starting point for my search.

Fenwick et al. (2008) defined the expected value of implementation as a function of the incremental net benefit of health technologies and the current and perfect level of implementation. Together with the expected value of information, this framework could be used to calculate the ceiling values of conducting further research (through the EVPI) and investing in implementation measures (through the EVPIM), and both through the *Expected Value of Perfection (EVP)*. Information and implementation were assumed to be independent; that is, further research evidence would not affect implementation levels. The authors outlined areas for further research that included: (1) relaxing the assumption of independence of implementation and information; (2) incorporating uncertainty around implementation levels in the analysis; (3) evaluating specific levels of implementation that are attainable with the use of implementation strategies, thereby allowing states other than current and perfect; (4) application of the framework in an ongoing appraisal. In addition to these, implementation is a dynamic process and the dynamic nature of implementation should also be considered when applying or developing the framework by Fenwick et al. (2008).

The objective of this review was to find out about any further developments and applications of the expected value of implementation and information framework developed by Fenwick et al. (2008).

2.5.2 Methods of value of implementation review

2.5.2.1 Search strategy

To cover both, scientific and grey literature, I performed a search in three search engines: 1. Medline, 2. Web of Science and 3. Google Scholar. The search was a citation search; that

is, I performed citation searches of the named journal article by Fenwick et al. (2008) in all three search engines. In the first phase, title and abstracts were reviewed, and the full texts of included articles were reviewed in the second phase.

2.5.2.2 Exclusion criteria

Excluded were all search results that were not an application or further development of the framework developed by Fenwick et al. (2008). Duplicates were also removed.

2.5.2.3 Data extraction and analysis

The approaches to apply the Fenwick et al. (2008) framework and to develop it further are presented in the results section in a tabular overview (which also includes the Fenwick et al. (2008) paper) and narrative. Furthermore, I examined whether the abovementioned identified areas for further research had been addressed since the publication of it to date (October 2015), leading to the following research questions:

1.) Has the framework developed by Fenwick et al. (2008) been applied in a practical example?

2.) Has the framework developed by Fenwick et al. (2008) been developed further in any way?

3.) Were implementation levels other than current and perfect considered?

4.) Was the effect of research evidence on implementation considered?

5.) Were implementation dynamics considered?

6.) Was uncertainty incorporated around the levels of implementation?

2.5.3 Results of value of implementation review

The three searches, which were updated in October 2015, resulted in a total of 115 hits, before exclusion criteria applied. Out of those, 20 hits came from the Medline search, 16 from the Web of Science search and 79 from the Google Scholar search. After 25 duplicates were removed, there were still 90 hits to be reviewed. After the title and abstract review, 70 were excluded based on exclusion criterion (A), as they were no applications or further

developments of the EVPI and EVPIM framework developed by Fenwick et al. (2008). One further hit was excluded because the link led to a Chinese university website that appeared to link back to the Fenwick et al. (2008) paper. Translation was therefore not considered worthwhile. Out of the remaining 19 hits, 11 were excluded based on the full text, leading to eight included publications. Also known to the author was another report, which was included as Entry 8 (see Appendix A.3). A further study that was only published online in November 2015 and known to the author was also added as Entry 11. The 10 included publications as well as the publication by Fenwick et al. (2008) are described in chronological order and are critically appraised in Appendix A.3.

The answers to the six research questions are summarised in the following:

1.) Has the framework developed by Fenwick et al. (2008) been applied in a practical example?

Based on the findings from this review, eight applications of the framework, additional to the original publication, have been identified. The identified publications have either calculated the EVPIM and EVP only (Hoomans et al., 2009c, Soeteman et al., 2011, Tuffaha et al., 2014, Revill et al., 2015) or have also produced the expected value of specific implementation measures (Hoomans et al., 2011, Faria et al., 2014, Whyte et al., 2014). One article (Willan and Eckermann, 2010) furthermore estimated the effect of information on implementation in two example case studies and another allowed for states other than perfect and current (Andronis and Barton, 2016).

2.) Has the framework developed by Fenwick et al. (2008) been developed further in any way?

Three publications provided extensions to the original framework. Willan and Eckermann (2010) presented a framework for estimating the value of information allowing for imperfect implementation, and by modelling the effect of research evidence on implementation as a function of the strength of evidence. This was done by employing a sliding step function for implementation, which depended upon the strength of evidence. The latter was represented by a certain threshold of the z-statistic; once it was surpassed, implementation would linearly increase until the improved implementation level was reached. The authors calculated the EVSI and demonstrated that the optimal research design could be found by maximising the expected net gain (that is, the EVSI compared to trial costs) associated with different designs. The authors found that the EVSI increased with

imperfect implementation. This was because research also has an effect on the implementation, thus reducing the expected opportunity loss incurred by the health care payer due to a proportion of patients being on a treatment that is expected to be less cost-effective. However, the dynamic nature of implementation was not considered and implementation was assumed to jump to the calculated level when evidence became available. The relationship between information and implementation was furthermore assumed to be linear between the current level of evidence and a defined future level of evidence at which implementation would reach the highest possible level.

The report by Walker et al. (2014) had the aim of providing a framework to allow for the evaluation of implementation initiatives in the England and Wales NHS. The framework was essentially the same as the one proposed by Fenwick et al. (2008) but the authors allowed for the following extensions: a) they presented a way of evaluating implementation initiatives that could target different subgroups of a population, accounting for different costs, benefits and utilisation rates; b) they allowed for multiple periods in the calculation of the value of implementation measures; c) they considered imperfect levels of implementation that would result from implementation measures using the 'expected value of actual implementation'. The authors let implementation levels differ in each period and costs of the implementation initiative could occur in different periods of time.

The framework by Walker et al. (2014) was applied in follow-up reports shown in Entry 7 and 8 in Appendix A.3, in a case study on novel anticoagulants in the prevention of stroke and systemic embolism (Faria et al., 2014) and another application to B-type natriuretic peptide (BNP) testing in diagnosing chronic heart failure (Whyte et al., 2014). The latter study is now published as a peer-reviewed journal article (Whyte et al., 2016).

Andronis and Barton (2016) extended the existing four-state framework by two further states: improved implementation and sample information. They then presented the 'implementation-adjusted' EVSI and claimed that the traditional EVSI would over-estimate the real value of research when implementation was imperfect. This is a simplification that does not recognise that the 'implementation-adjusted EVSI' is, in fact, nothing but the value of a research study treated as an implementation measure, thereby disregarding the effect of research on decision uncertainty altogether. This is because the 'implementation-adjusted EVSI' was calculated omitting the maximisation of expected net benefits under current information, and the maximisation of net benefits in each simulation for the with research scenario, and instead uses an uptake-weighted average net benefit in both cases. The effectof research on uncertainty is thereby ignored. What is truly calculated is the expected value of implementation of the research activity. If the authors had explored a sufficiently large range

of implementation levels, including low ones, they would have found that their 'implementation-adjusted' EVSI can, in fact, be larger than the EVSI. This is intuitive because EVSI and 'implementation-adjusted' EVSI measure different things. One measures the potential reduction in decision uncertainty that a research study can achieve and the other measures the potential reduction in the expected value of implementation that is attainable with the same research study. More will be said on this in Chapter 4.

3.) Were implementation levels other than current and perfect considered?

Implementation levels other than 'current' and 'perfect' were allowed in one series of studies (Faria et al., 2014, Walker et al., 2014, Whyte et al., 2014) and another publication by Willan and Eckermann (2010). Estimates for these states were based on assumption or derived by quantitative extrapolation methods described in Question 5.) below. Another article (Andronis and Barton, 2016) used static implementation estimates that were *'elicited from experts'* but no uncertainty was reported. In this study, other scenarios were also tested in sensitivity analysis based on assumptions.

4.) Was the effect of research evidence on implementation considered?

Only two studies considered the relationship between the strength of research evidence and implementation levels (Willan and Eckermann, 2010, Andronis and Barton, 2016) as described in Question 2.). Problems with the Willan and Eckermann (2010) approach may lie in the fact that the strength of research evidence reflects uncertainty surrounding a specific parameter, rather than decision uncertainty. This may or may not be appropriate in reflecting a trust's or clinician's thought process when making the adoption decision.

The problems with the Andronis and Barton (2016) approach relate to the authors' interpretation of the 'implementation-adjusted' EVSI as described under Question 2.). A further problem relates to the way they obtained the before and after research implementation estimates. Experts provided point estimates of 50% for the current level of implementation and 75% for the with research implementation and a linear improvement in implementation of 5 percentage points per year for the cost-effective technology was assumed to occur in the following years.

5.) Were implementation dynamics considered?

In Faria et al. (2014), linear and polynomial regression methods were used to extrapolate future implementation from available data over a three year period, but depending on the method, the resulting implementation levels after some periods of time varied widely. The authors concluded that it was difficult to assess which model can predict future utilisation accurately. The wide variation of results may have been caused by an arbitrary choice of methods that ignored well established methods of modelling diffusion, as reported in a variety of studies, including the 25-year review of diffusion modelling methods by Meade and Islam (2006).

In the paper by Whyte et al. (2016), an exponential model was used to obtain future implementation levels that followed an s-shape, acknowledging diffusion theory (Rogers, 2003). However, evidence was only available on two periods of utilisation, and it appeared as though these were not the first two periods after technology introduction. This means that the inflection point may not have been accurately captured as data on all periods since technology introduction are required for sigmoid curves to represent diffusion accurately (Meade and Islam, 2006). Furthermore, the maximum number of attainable adoptions was implicitly set at a hundred percent of the eligible patient population, which may not be realistic, and possibly a result of ignoring the first periods of diffusion. Both of these issues may result in arbitrary effects on the fitted diffusion curves, with the second effect possibly being the over-estimation of uptake that can potentially be achieved.

Andronis and Barton (2016) used implementation estimates that changed linearly in the future. They used a sigmoid curve for implementation in a sensitivity analysis, which they stated was inspired by Rogers' diffusion theory (Rogers, 2003) but they did not provide any further detail on it. All three applications did not use established diffusion models for which the fit with real world data has been shown. These are examined further in Chapter 3.

6.) Was uncertainty incorporated around the levels of implementation?

None of the included studies incorporated uncertainty in their implementation estimates.

2.5.4 Discussion of value of implementation review

The expected value of implementation review showed that the value of implementation was assessed in a small number of applications only. The unified framework originally developed by Fenwick et al. (2008) was extended further by: 1. allowing for subgroups, 2.

allowing the strength of research evidence to affect implementation, 3. allowing implementation levels other than 'current' or 'perfect' and finally 4. allowing implementation to change over future time periods. Whilst three included papers explored future diffusion with a nod to diffusion theory, none of them used established diffusion curves for which there is evidence on the fit with real world data. This is important because not all s-shaped extrapolations have demonstrable forecasting ability (Meade and Islam, 2006). None estimated the uncertainty associated with their estimates of diffusion.

These findings demonstrate that more research is needed in the area of obtaining implementation estimates. More concretely, it was shown that implementation estimates are needed for: 1. the 'counterfactual' development of implementation levels, when no implementation initiatives are pursued; 2. the 'potential' with implementation initiatives or with research implementation levels; and 3. implementation levels for each future period in both of these cases.

Two particular weaknesses in the use of implementation dynamics were seen in the method used for extrapolating future implementation: 1.) in the study by Faria et al. (2014), the regression method used significantly influenced the maximum level of implementation that could potentially be achieved. 2.) The data used for the observed implementation levels as well as for the effect size of implementation measures were highly uncertain in the study by Whyte et al. (2014), and not all periods since technology launch had been incorporated in the analysis.

It may be desirable from a policy perspective to identify technologies or policies with potentially low implementation early and to assess the value of implementation measures before they are introduced in the health system. This may be particularly relevant when research studies may have an effect on implementation, as research is more likely to be conducted before technology introduction. Performing such analysis before technology introduction could thus enable policy-makers to design research studies with the effect on implementation in mind, similar to a CTA approach used in the Netherlands. None of the included publications had attempted to estimate implementation levels before technology introduction. Such analysis is likely to be associated with further methodological difficulties of obtaining implementation estimates and identifying those technologies with potentially low implementation. Furthermore, the effect of research evidence on implementation should be made more specific than the sliding-step relationship of the z-static and future implementation levels employed by Willan and Eckermann (2010).

The strength of this review was its targeted and specific methodology. Only the addition of Google Scholar as another database meant that the number of search results went up to more than a hundred – without Google Scholar there would have been only 36 hits before removing duplicates. Including Google Scholar proved important: three of the included publications were only found by Google Scholar and two of those were especially relevant in that they had developed the original framework further.

The limitation of this review was the same as the strength. Being very specific and limited to only three databases meant that other publications studying the value of implementation may have been missed, for example because of the databases not accessing all relevant journals. This was not considered a sufficiently large problem to warrant the extension of this review to other databases. Being a citation search also meant that studies on a similar topic that did not cite Fenwick et al. (2008) would have been missed.

In conclusion, this review showed continuous interest in the expected value of implementation from the time of publication of the Fenwick et al. (2008) framework, which was applied and extended in 11 studies until November 2015. It also highlighted the need for further research, in particular around the estimation of implementation dynamics, potentially before technology introduction, and the relationship between research and implementation.

2.6 Summary

The reviews presented in this chapter found that implementation estimates were not commonly used in economic evaluation of health technologies. The uses of implementation estimates were limited to expected value of implementation analysis, population estimates of value of information, costs or effects, and the modelling of infectious diseases. When implementation estimates were used, the approaches of obtaining them were basic: most studies with few exceptions (Faria et al., 2014, Whyte et al., 2014) did not consider changing implementation over time and estimates were mostly based on reference data or assumption.

These findings provided guidance in the selection of the areas for further research within this thesis. Most importantly, it became clear that further study of methods to obtain estimates of implementation over time was needed. I also identified some scope for further research and development of the expected value of implementation framework. Furthermore, price changes precipitated by the reimbursement decision and subsequent changes in implementation had not been included in economic evaluation.

The research topic was approached from different angles, using different types of reviews (scoping and citation searches) and different key publications as starting points. This ensured that a variety of research studies were included. The citation searches were narrowly defined but I was confident to have found the most relevant studies in the respective areas. The scoping review highlighted the challenges commonly associated with methodological research: a large number of hits was obtained of which only a fraction was relevant. It did help with the identification of areas for further research but the lack of comprehensiveness meant that not all the studies using implementation estimates were found. However, I was confident that no important methodological study was missed after having performed the other searches and reviews. Multiple presentations of this work in front of different academic and policy audiences did not result in any further studies being highlighted to me.

Further research is needed in obtaining an overview of actual health technology diffusion data in England and to review potential methods to model health technology diffusion. Model frameworks are required that can facilitate the incorporation of diffusion estimates in value of information and implementation analysis and the modelling of the diffusion-price relationship through experience curves.

In conclusion, there is scope for further research on the incorporation of diffusion estimates in health economic methods.

CHAPTER 3. REVIEWS OF METHODS TO ESTIMATE HEALTH TECHNOLOGY DIFFUSION PRIOR TO TECHNOLOGY INTRODUCTION

3.1 Background

The previous chapter established that diffusion estimates are relevant for economic evaluation of health technologies. As was highlighted in the introductory chapter, implementation of new technologies follows a gradual growth curve over time. This is explained through diffusion theory. The natural shape of health technology diffusion, however, is unknown. The fact that HTAs are commonly performed prior to technology introduction presents an additional methodological challenge in the estimation of diffusion for the technology to be assessed.

The aim of this chapter was to identify a method of obtaining probabilistic estimates of new health technology diffusion with no existing data on uptake. This method would then be used to obtain diffusion estimates for the case study technology (Chapter 6) (see this chapter's place in the thesis in Chapter 1, Figure 1.2). In Section 3.2, I explore available evidence on health technology diffusion data in the UK. The objective of this was to provide the information needed for generating a synthesised diffusion curve of health technologies in the UK, potentially controlling for influencing factors such as the effect of NICE guidance. Such synthesised curves could then be used for future reference, potentially with some adaptation, to predict new health technology diffusion. In Section 3.3, I explore existing sales forecasting methods and diffusion models that could help predict diffusion of individual health technologies. The objective was to identify a method to generate probabilistic estimates of diffusion of individual health technologies when no data is available. For this, I reviewed existing diffusion models and sales forecasting methods with the ability to predict diffusion curves when no data is available.

3.2 Review of health technology diffusion in the United Kingdom

3.2.1 Objective of UK health technology diffusion review

This literature search sought to identify publications with empirical estimates of diffusion curves of health technologies in the UK health care setting that could potentially enable evidence synthesis on health technology diffusion in the UK. I chose the UK instead of England for this review, to avoid exclusion of evidence that could potentially be adjusted to reflect England.

3.2.2 Methods of UK health technology diffusion review

3.2.2.1 Search strategy

The search strategy followed three different branches. These branches were each targeted reviews to identify additional relevant publications and are described in Sections 3.2.2.2 to 3.2.2.4. The searches were performed in January 2013.

3.2.2.2 The ERNIE database

The Evaluation and Review of NICE Implementation Evidence (ERNIE) database was searched for so-called NICE implementation uptake reports, which empirically estimate the uptake of NICE technology appraisals, clinical guidelines and public health initiatives. The ERNIE database also lists publications on uptake from external sources, referring to external publications that were related to health technologies previously assessed by NICE. There were references for each of the following categories: cancer service guidance, clinical guidelines, interventional procedures, public health guidance and technology appraisals. I deemed the technology appraisals and interventional procedures streams as most relevant for the research question. References in these two categories were therefore searched. There were five interventional procedures and 154 technology appraisals with a varying number of between one to ten external references for each of them. References for all five interventional procedures and all 154 technology appraisals were searched.

A few external references contained information on various health technologies. This resulted in those key references being listed as external references for more than one technology appraisal. Those were marked as ‘duplicates’. In the case of irrelevant publications the reason for their irrelevance was given, whether they were duplicates or not, instead of marking them as duplicates first. This followed from the structure of the database that did not allow exporting of references.

3.2.2.3 Medline

The search strategy for Medline was developed following an initial search for key publications and after identifying suitable search terms from their abstracts and keywords.

Keywords for search Medline 1 were:

(Diffusion OR Uptake OR Adoption) AND (adj3 Curve OR adj2 Data) AND ('Diffusion of Innovation') AND (geographical filter UK)

The subject heading 'Diffusion of Innovation' was dropped for one search variation to explore whether more results would be obtained (Search Medline 2). The geographical filter was dropped for another variation of the search (Medline 3). Results were limited to publications in English language.

3.2.2.4 Restrictive Google Scholar search

A restrictive Google Scholar search was conducted using the following search terms:

(Diffusion OR Uptake) AND curve AND health

Only English language publications were retrieved and the search was limited from January 1991 to January 2013. The rationale for this was to limit the search to recent and therefore more comparable data. The first one hundred results were screened for relevant publications that might not have been picked up by the other searches, especially grey literature.

3.2.2.5 Inclusion and exclusion criteria

Included were publications with diffusion data in the UK health care setting. Excluded were publications using data from outside the UK or with a data series of less than two years. Publications on technologies that could not qualify as health technologies or were grouped rather than individual technologies were excluded. Further publications were excluded when the terms diffusion or uptake did not refer to diffusion of health technologies but instead to diffusion of other phenomena. Publications were also excluded when they were secondary publications of data for which the primary publication was available.

3.2.2.6 Data extraction

The results of the searches were described in narrative. The resulting diffusion curves were categorised in terms of conditions that were grouped as cancer, acute conditions, chronic conditions, conditions that can be chronic and acute and others. Data for the condition, the technology, the measure of diffusion or proxies used, the data source, the length of time for which diffusion data was available and the first year diffusion data was extracted into tabular overviews.

3.2.3 Results of UK health technology diffusion review

3.2.3.1 ERNIE database search results

This search identified 31 NICE implementation uptake reports, 16 of which referred to technology appraisals, 14 to clinical guidelines and one to a public health initiative. Out of the 16 reports on technology appraisals, 14 included data points on the technologies' uptake. Two were excluded for containing grouped data on several medications. Nine out of the 14 uptake reports concerning clinical guidelines yielded data points on diffusion. Five were excluded: one of them for not containing diffusion data; the other was a duplicate that was included in the technology appraisals uptake report and three contained grouped data for different technologies. One uptake report assessing a public health initiative was included in this review. In total, 24 uptake reports were thus included.

From the search of ERNIE external references, five interventional procedures with six external references were listed. None of them contained data on utilisation over time and so were excluded from the review. For the 154 screened technology appraisals, a total of 399 external references were listed. Out of those, 165 did not contain diffusion curves, 219 were duplicates of already included studies and one provided prescription volume data but only on drug classes instead of individual health technologies. Fourteen publications were included for full text review. Out of those, two were excluded because they referred to the same diffusion data as other publications. One was the innovation scorecard which did not yield diffusion curves of at least two years. One could not be found online. Ten publications were finally included.

3.2.3.2 Medline search results

Medline search 1 yielded four records. Two of those were included in the final results based on the inclusion criteria. One was excluded because the data used was from Florida, another because it provided data for less than two years. Two of the included publications referred to the same technology and to the same dataset and one of them was therefore excluded. One publication with diffusion curves of different technologies in the UK health care setting was included.

Medline search 2 yielded 16 records. None of them were included in the final results. As expected, the four records from Medline 1 came up again and were excluded; the 12 additional publications did not use UK data.

Medline search 3 yielded 169 records. After the abstract search, 139 publications were excluded based on their different understanding of the term diffusion which did not refer to diffusion of health technologies but, for instance, to the diffusion of tumour tissue. Ten publications were excluded because they did not contain data to inform a diffusion curve and two studies were excluded because their data were not from the UK. Fourteen potentially relevant publications were identified. As expected, the four results from search Medline 1 were listed but excluded as duplicates. After the full text review, three publications were included based on the inclusion criteria. The remainder were excluded as no diffusion data over time was provided (10) or because of grouped data (1).

3.2.3.3 Restrictive Google Scholar search results

The Google Scholar search yielded 47,700 records and the first 100 records, sorted by relevance, were searched. In 68 records, diffusion or uptake was used to describe diffusion other than that of health technologies. Nineteen records did not contain empirical diffusion curves, seven records did not come from the UK and three were not health technologies. Two had been identified through the Medline searches and were excluded. Therefore, one publication was found to be relevant and it was included after full text review. No further searching was undertaken using Google Scholar due to this poor hit rate and the availability of data from the other searches.

An overview of the number of publications retrieved by the searches and included after review can be found in Table 3.1.

Table 3.1 Overview of search results

	ERNIE implementation uptake reports	ERNIE external references	Medline 1	Medline 2	Medline 3	Google Scholar	Other sources	**Sum**
Records	31	405	4	16	169	100	1	**726**
Recommended for full text review	31	14	2	0	14	1	1	**63**
Included	**24**	**10**	**1**	**0**	**3**	**1**	**1**	**40**

3.2.3.4 Summary of findings

A total of 164 diffusion curves were reported within the 40 included studies. Diffusion curves related to 52 therapeutic areas or conditions. Conditions were grouped as cancer (10), acute conditions (9), chronic conditions (17), conditions that can be chronic and acute (8) and others (8). Others included, for example, smoking cessation programmes, obesity and prevention of diseases.

Diffusion was shown for a total of 127 technologies, 99 of which were drugs and 28 were other treatments or technologies. Those other treatments or technologies encompassed surgical technologies such as laparoscopic surgery, therapies such as computerized cognitive behavioural therapy, and devices such as Magnetic Resonance Imaging (MRI) scanning. Diffusion data was mainly presented as absolute numbers of prescriptions or prescription costs. Some publications used number of patients or number of milligrams or hospital episodes. Most of the data was presented as absolute numbers; less frequently relative to the incidence or as the share over other alternative technologies.

Different data sources informed the diffusion data in the included publications. Data were primarily obtained from the IMS Health Hospital Pharmacy Audit Index database, the electronic prescribing analysis tool (ePACT) and the prescription cost analysis (PCA) system. Less frequently, data were obtained from the Hospital Episode Statistics (HES) national data warehouse, surveys and General Practitioner (GP) records as well as specialised databases such as the British Pacing and Electrophysiology Group (BPEG) Register.

The available time series varied in length and most publications presented data on a limited number of periods. The average length of time series data was 6.5 years. The average length of data available without guidelines was 2.75 years. The length of time series data

was potentially too short to observe sigmoid curves in most cases although visual inspection resulted in a few sigmoid candidates (for instance, oxaliplatin and capecitabine for colon cancer; percutaneous coronary intervention with drug eluting stents in angioplasty; or donepezil in Alzheimer's disease). Characteristics of data on diffusion curves from the included publications – grouped as cancer, acute conditions, chronic conditions, conditions that can be chronic and acute and others – are provided in summary tables within Appendix B. As well as variation between studies in terms of conditions, diffusion measure, data sources and length of the time series, it was clear that contextual factors played a strong role in diffusion patterns. Consequently, no attempt was made at meta-analysing or summarising the identified curves.

3.2.3.5 How results of this review could inform further analysis

The presented search resulted in identification of relevant publications and extracted data that could be of potential value for estimating health technology diffusion in this thesis. The identified reports and articles that presented data on health technology diffusion in the UK did not include data on all health technologies currently in use in the UK. It was not possible to identify the selection criteria for health technologies for which diffusion data was reported. The data may therefore not be representative of the true technology diffusion in the NHS. Other issues relate to:

a) Availability of data

The exact data was mostly unavailable from the identified reports which mainly visualised diffusion in graphs. However, reasonably accurate data could be generated through digitising graphs using purpose-built software (e.g. Grafula).

b) Range of technologies covered

A wide range of technologies were covered by the literature. The proposed categorisation of technologies was to group them by conditions. In addition, drugs and other health technologies could be separated. There was data on only 28 technologies falling in the category of 'other health technologies'. These could further be grouped as: types of surgery (7), diagnostics (4), devices (9) and others (8).

c) Heterogeneity

Even within technology / disease groupings, diffusion curves appeared extremely diverse, with some stagnating at a very low level and some going through exponential growth and a following decline. None of the diffusion curves formed a straight horizontal line from the date of technology introduction. No explanation was provided for the different shapes of the curves. The only piece of information that was available for many of the included diffusion curves was the issue date of NICE guidance. Based on visual inspection, it could not be observed that the issue of NICE guidance resulted in a homogeneous change in the examined diffusion curves. Due to the lack of information on explanatory factors, meta-analysing the extracted data would have likely produced meaningless results. With knowledge of these explanatory factors, diffusion curves could be grouped and aggregated more sensibly, which could then enable meta regression analysis. However, this would require in-depth research for each technology on information that is not available from this review. None of the included studies examined drivers of diffusion in an attempt to explain heterogeneity.

d) Length of time series

Ideally, time series would cover the entire life-span of a technology, or at least include the inflection point of the curve. However, none of the present data sets fulfil the former, and based on visual inspection it is difficult to assess whether the latter is given. The stage a technology is at in the product lifecycle is an important explanatory factor that would need to be identified for each set of diffusion data if they were to be analysed. However, length of time series is also important as it impacts on the accuracy of longer-term estimates. If the decision was made to only take diffusion data with more than three years of observed data into account, then 133 diffusion curves were still available. With greater than seven years of diffusion data, the number of curves dropped to 52 and greater than nine years of data decreased the number of diffusion curves to 25. The mean length varied over the different categories of conditions, with diseases that can be both acute and chronic having the longest diffusion curves (mean of nine years) and the cancer category having the shortest (mean of five years).

As the issue of NICE guidance could have an impact on ‘natural’ diffusion, it would be of particular interest to analyse this, conditional on sufficient data points being available. On

average, diffusion curves covered two to three years before guidance was issued. Only analysing diffusion curves of three years before guidance was issued reduced their number to 32. Only six diffusion curves were left when diffusion curves with more than seven years before guidance were included.

e) Specificity of data

There were a few studies that showed prescription data of, for instance, proton pump inhibitors or statins, without specifying the condition. These were included in the present results but would need to be excluded for further analysis as they would not be suitable for assessing the diffusion of a health technology for a specific condition.

f) Output variable

Some of the included publications used prescription volume data as a measure of diffusion. Others used prescription costs as a proxy, which seemed reasonable under the conditions that prices did not change over time and that they were split by indication. Further studies used the proportion of usage of a technology relative to the eligible patient population. The latter appeared to be the most sensible option for aggregation of data across conditions with different population sizes. A diffusion measure relative to the total eligible population would, however, require further data on the expected patient population over time or in the case of costs, the expected total treatment costs over time.

3.2.4 Discussion of UK health technology diffusion review

This review showed that diffusion data on health technologies in the UK were available. The above-mentioned heterogeneity and the short length of time over which data were available meant that a meta-analysis would not yield sensible and usable results. While heterogeneity could potentially be addressed by further qualitative or quantitative analysis to better explain the impact of different factors influencing diffusion, the number of available diffusion curves with sufficient time periods of data available was likely too small to enable meaningful meta-analysis. This review did not yield the qualitative information necessary for such analysis, for example information on entry of competitors and other potential diffusion determinants.

Another option of generating a diffusion curve for a particular health technology would be to choose an analogous technology from this catalogue and use its diffusion curve, possibly with some adaptation. The caveat is that the defining features of an 'analogous' technology are not entirely clear in the context of health technology diffusion. In other studies, such analogous technologies are typically very narrowly defined as those in the same product category and with other similarities (Jiang et al., 2006). Those similarities could include the environment within which the product will be launched, the behaviours of buyers, marketing strategy and the nature of the innovation (Goodwin et al., 2014).

The implication for the further development of this thesis was that using meta-analysed diffusion data was not a feasible option and that picking an analogous technology and adjusting its diffusion to match the context in question may be challenging. Other methods are therefore required to obtain diffusion estimates. With NICE publishing the implementation uptake report series, much more diffusion data is going to become available in the future, which could potentially enable the future use of analogous technologies or meta-analysis.

The strengths of this study relate to the number of databases explored. This led to a broad range of health technology diffusion data found. Limitations were the restriction on the Google Scholar search to only one hundred hits and the exclusion of public health guidance, cancer service guidance and clinical guidelines databases for searching external references. However, these searches were not expected to reveal sufficient additional diffusion data to change the conclusion of this review.

In conclusion, the health technology diffusion evidence available for the UK is not mature enough to perform meta-analysis from empirical data alone and other methods of estimating diffusion are required.

3.3 Review of methods for predicting diffusion prior to launch

3.3.1 Background to predicting diffusion methods review

With observed evidence of diffusion curves not being mature enough to perform evidence synthesis of empirical data alone (as demonstrated in Section 3.2 of this chapter), another estimation method for health technology diffusion was needed. This section therefore sought to identify methods to predict diffusion. One starting point was having

informal conversations with two representatives of one medical device and one pharmaceutical manufacturer, with the objective to identify methods of forecasting or a body of literature. These did not identify formal forecasting methods but suggested that the prediction of future sales entailed consideration of multiple variables such as the number of physicians covered by sales representatives per year and the likely market share that could be achieved.

Another starting point was to explore methods commonly used for sales forecasting in pharmaceutical or health technology companies. A scoping review was performed in the databases Business Source Premier via EBSCO, Medline via OvidSP and Google Scholar on search terms relating to sales forecasting and diffusion of health technologies. This scoping review produced results on a diffusion model called the Bass model of new product growth (Bass, 1969, Sillup, 1992, Meade and Islam, 2006, Bass et al., 2001).

For lack of knowledge of formal forecasting methods other than the Bass model of new product growth, I decided to do further research on this established model. This choice was supported by the fact that values of model parameters were available for many different technologies including health technologies, and its fit with real data was demonstrated in several publications (these are discussed in the sections below). The Bass model of new product growth (Bass, 1969), developed shortly after Rogers' theory of diffusion (Rogers, 2003) described in Chapter 1, used a logistic model to reflect the s-shape of cumulative technology diffusion, with parameters adapted to reflect the degree of innovation and imitation as well as the overall attainable number of adoptions:

$$n(t) = p(m - N_\tau) + \frac{q}{m} N_\tau (m - N_\tau) \tag{3.1}$$

where $n(t)$ is the number of new adoptions in period *t*, with $n(t) \geq 0, t > 0$, p the coefficient of innovation, and q the coefficient of imitation, with $\frac{q}{p} > 1$ to ensure the s-shape (Meade and Islam, 2006), m the total number of attainable adoptions with $m > 0$, N_τ the cumulative number of adoptions up to τ with $\tau = t - 1$. The coefficient of innovation (p) can be explained as the speed of adoption in initial periods that is largely independent of peer influence. The coefficient of imitation (q) is the speed of adoption in later periods at which peer influence is the main driver.

3.3.2 Methods for predicting diffusion methods review

3.3.2.1 Search strategy

I performed a state-of-the-art review (Grant and Booth, 2009) to identify papers on methods to predict diffusion without any available data. It is worth mentioning here that a review of pre-launch forecasting methods was published in 2014 (Goodwin et al., 2014) that was found when updating this review in December 2015. This is especially relevant to this thesis as it focusses on pre-launch forecasting methods and includes methods other than diffusion models. This is going to be discussed together with other findings in Section 3.3.3.

A state-of-the-art review is a review that prioritises most current methods to point out areas for further research. From the scoping review described in Section 3.3.1, one study was identified that predicted diffusion with no data available: a paper by Bass et al. (2001). Based on a systematic review by Meade and Islam (2006) that had the objective to identify extensions and adaptations to diffusion models (Meade and Islam, 2006), the Bass et al. (2001) paper was the only one predicting diffusion with no data available up to the time of their review. I therefore used the Bass et al. (2001) paper to perform a citation search in Web of Science, with the objective to identify any newer papers published since, in which diffusion was predicted prior to technology launch. This search was first performed in May 2013 and updated in December 2015.

3.3.2.2 Inclusion and exclusion criteria

The titles and abstracts of identified publications were screened and papers were included for full text review based on the inclusion and exclusion criteria below. These criteria were applied again in the full text review. Papers were included when methods were described for predicting diffusion without any data available. Papers were excluded when (a) no methods for predicting diffusion were described, (b) there was some data available and (c) the methods were already described in the original paper that was included.

3.3.2.3 Data extraction

Included papers as well as the paper by Bass et al. (2001), the starting point to this search, were summarised in terms of their objectives, methods used and limitations in tabular format. Important developments in methods for predicting diffusion with no data available were additionally described in narrative. In addition to the review results, the critical review by Goodwin et al. (2014) was summarised in order to provide an overview of forecasting methods in general.

3.3.3 Results of predicting diffusion methods review

3.3.3.1 Included publications

The citation search in Web of Science resulted in 18 citations of the paper by Bass et al. (2001). Ten of them were excluded based on the title and abstract screening, four because the papers did not describe methods for predicting diffusion (a), five because the predictions were based on some available data (b) and one because the original paper referenced was already included (c) (the review by Meade and Islam (2006) that was identified before this review, as described above). Three more papers were excluded during the full text screening phase, two of which based on criterion (b) and one based on criterion (a). The full text review therefore resulted in five included papers. The review by Goodwin et al. (2014) was not included in the discussion of this review but was discussed separately (Section 3.3.3.3).

3.3.3.2 Findings from the search and review on predicting diffusion

Five out of six of the papers included in this review used a method called 'guessing by analogy' to predict diffusion of a new technology prior to technology introduction, when no data were available (Appendix B, Table B.7). The Bass et al. (2001) paper predicted diffusion of a new television technology by analogy and expert judgement was used to identify the most appropriate analogous product. The authors stated that it was difficult to choose the right analogy because little was known on how analogies should be chosen. Jiang et al. (2006) expanded on this stating that guessing by analogy is of limited applicability when few similar products are available, especially because it is difficult to find technologies sufficiently similar to result in similar diffusion patterns. Furthermore, data of analogous technologies are often truncated on either side, that is, there usually is data shortage on the beginning phase of diffusion or on the later periods (Jiang et al., 2006). This introduces bias (Jiang et al., 2006). Diffusion data may also be more likely to be published for successful technologies causing sample selection bias. Finally, elapsed time since the analogous technology was introduced may have brought about other exogenous factors that may influence diffusion patterns (Kim et al., 2013).

An extension to commonly used guessing by analogy methods was presented in the paper by Lee et al. (2014). In their study of consumer electronics, product attributes of a large number of technologies in the same product category were analysed for their effects on product diffusion. This was achieved using statistical and machine learning-based

approaches. For this, the authors created databases of product attributes that were informed by expert judgement, and of diffusion characteristics that were achieved through non-linear least squares regression analysis of at least 12 years' worth of available adoption data points for 80 technologies. The product attributes were developed based on a literature review and discussions with senior marketing managers. Four of the 17 identified attributes were valued on nominal scales and the others were valued on five-point Likert scales from very low to very high. All of the attributes were valued for each of the included technologies by industry experts. Among the various regression methods employed, multivariate linear regression performed best in terms of mean absolute error.

While the method proposed by Lee et al. (2014) is a step in the direction of obtaining better diffusion estimates prior to launch based on analogous technologies, it is unlikely to be reproducible in health technologies. This is because a large database of similar products is required to establish the relationship between product attributes and diffusion characteristics. The degree to which technologies are similar is important: in non-health technologies, analogies were typically used from the same technology class, when there were no long lags between the launch dates of the respective technologies (Goodwin et al., 2014). This is much more likely to happen in consumer markets in which the protection of intellectual property plays a lesser role than in health care. Lee et al. (2014) used a database of 80 technologies. In health technologies, it is improbable that such large numbers of technologies in the same therapeutic area and with similar product characteristics can be found. Furthermore, to be able to produce relatively accurate diffusion estimates, the available adoption data need to encompass the peak of per period adoptions. For this reason, Lee et al. (2014) only included those products with more than 12 years' worth of adoption data. Considering that the review presented in Section 3.2 found only 25 diffusion curves that had more than nine years of data available, it is unlikely that the numbers needed for such analyses can be generated in the health technology sector at present.

An example of a subjective judgement method to product demand forecasting was presented in the paper by Kim et al. (2013). The authors proposed a survey method that asked for seven items associated with important data on diffusion. The seven items were: quantity in first period, time and cumulative quantity at take-off, time and cumulative quantity at peak, time and cumulative quantity at stagnant demand. The authors chose the time of stagnant demand over the quantity of market potential, because the latter was deemed difficult to estimate, being the asymptomatic value of demand over infinite time. These quantities could be converted to estimate the parameters of the logistic model by means of algebraic transformation and local searches.

This survey method stands in contrast to other studies that proposed using three questions on peak time, peak rate and market potential (Mahajan and Sharma, 1986) and sum of Bass model parameters *p* and *q*, initial demand and market potential (Lawrence and Lawton, 1981). The rationale for expanding the number of questions was to minimise the chance of erroneous estimates of one or more of the parameters. The authors considered such erroneous estimates more likely due to the vague nature of questions asked in those two studies. A limitation of this study was that the survey only delivered deterministic diffusion data and did not allow for the respondents' uncertainty to be expressed. If this was to be addressed by using formal elicitation of expert opinion, then the number of questions asked may be too great to provide meaningful results and fit into tight schedules of experts. The authors used a logistic model instead of the Bass model, owing to its relative simplicity. They did, however, acknowledge that the Bass model is superior in its ability to accurately forecast, especially in technologies for which the demand pattern is asymmetrical about the peak of per period adoptions. This study highlighted the difficulty in finding a good balance of information that is easy to elicit because it is intuitive, and information that can be transformed into diffusion model parameters.

3.3.3.3 Summary of Goodwin critical review of new product forecasting methods

The paper by Goodwin et al. (2014) contains a critical review of different methods of pre-launch forecasting. While this paper does not describe any prediction method based on diffusion models that were not included in this review already, it is a valuable addition in that it discusses forecasting methods other than those based on diffusion models, a body of literature that I had not found before. I therefore briefly summarise the content of this paper:

Goodwin et al. (2014) stated that the ideal scenario forecast included a probability distribution for adoption levels at each time period and that this had possibly never been accomplished. Instead, forecasting methods used included (i) management judgement, (ii) the analysis of judgements by potential customers and (iii) formal models of the diffusion process. Management judgement (i) was found to be elicited from individuals or groups, with aggregation methods typically spanning simple averages and Delphi methods, rather than more sophisticated aggregation methods such as behavioural aggregation (O'Hagan et al., 2006). Nothing in this review indicated that this method had ever included formal elicitation of expert opinion and diffusion models were also not used.

For method (ii), potential customers were typically asked about their likelihood of purchasing the new technology, either in terms of binary outcomes or on five-point or 11-

point scales. Other methods such as utility-based conjoint analyses or discrete choice experiments can help in establishing the probability of adoption when utility of the new technology exceeds that of alternatives. However, customer-based approaches are less likely to be applicable in the medical sector in which purchasing decisions are likely made by groups of people in a clinical commissioning group.

For approach (iii), Goodwin et al. (2014) described in detail only the Bass model of diffusion and stated that it was based on a firm theoretical rationale. The authors mentioned other possible models such as the Gompertz, logistic and Weibull functions as alternatives. The authors also reported alternatives, namely system dynamics models and agent-based modelling, but also highlighted that these had not been widely used and therefore lacked data on effectiveness in forecasting. The paper concluded with stating that formal diffusion models should be at the core of new product forecasting and were preferred to unstructured judgements.

3.3.4 Discussion of predicting diffusion methods review

This state-of-the-art review found that there are two types of methodologically and theoretically validated methods for predicting diffusion prior to launch: guessing by analogy and expert judgement on quantities informing diffusion models. Both methods have their limitations. Guessing by analogy is unlikely to deliver accurate forecasts in health technologies because of the limited number of appropriate analogies, even in its more sophisticated form that was proposed by Lee et al. (2014). The problem with expert judgement has been that only deterministic judgements have been elicited so far and that transforming intuitive quantities into diffusion model parameters can be tricky. There is therefore scope for developing expert judgement methods further to elicit probabilistic estimates and try and overcome the issue of transforming elicited quantities into model parameters.

A limitation of this search was the restriction to Web of Science as the only database searched. Diffusion literature is notoriously difficult to search for. Web of Science had previously been shown to include the most relevant search hits in a review described in Chapter 2.5 and had the feature of conducting citation searches. With the focus on peer-reviewed methods, Google Scholar was not considered relevant for this search. A quick browse through the 88 hits when performing the same citation search on Google Scholar suggested that the inclusion of Google Scholar as a database would not lead to the addition of publications with additional information. Furthermore, the critical review of pre-launch

forecasting by Goodwin et al. (2014) confirmed that no formal diffusion model estimation methods were missed.

In conclusion, expert judgement methods based on formal diffusion models appear to be the most appropriate method of predicting health technology diffusion when no data are available. This was because one alternative, guessing by analogy, is associated with the following problems when applied in the context of health technologies: a) the fact that there typically are few comparators in health technologies; b) that vast heterogeneity has been observed in health technologies, even those in the same class; c) the difficulty in finding truly analogous technologies in general; d) that large numbers of analogous technologies would be required to perform a more sophisticated guessing by analogy study in the spirit of Lee et al. (2014); e) even fewer analogies are available when stipulating that available time series data include the inflection point; and f) the introduction of bias when using immature diffusion estimates. Expert judgement methods are associated with problems of a) not using established diffusion models for which evidence of superior forecasting accuracy was available; b) the introduction of bias that is caused by unstructured estimation of parameters; and c) the use of vague questions. When using expert judgement methods based on formal diffusion models, there is also scope for further development in obtaining probabilistic estimates, for example through formal expert elicitation methods.

In conclusion, scope for developing a better method for pre-launch forecasting lies in the area of an elicitation of expert beliefs about diffusion that uses the structure of an established diffusion model.

3.4 Identifying the most appropriate diffusion model for this thesis

Having decided that elicitation of expert opinion will be used on the structure of an existing diffusion model, I will give a brief overview of different diffusion models and then more specifically the Bass model of diffusion, as well as extensions and adaptations that may be relevant to this thesis.

3.4.1 Use of the Bass model compared with alternatives

In their 25-year review of diffusion modelling, Meade and Islam (2006) describe a number of different diffusion models, the so-called epidemic models, that include the Bass model, the generalised Bass model, different logistic models, exponential models, the Gompertz model and Weibull models, amongst others. The authors also describe the use of individual level models, often simply called probit models, as reported by Geroski (2000), amongst others. Probit models provide the advantage of incorporating heterogeneity between adopting institutions while maintaining the ability to generate s-shaped curves (Geroski, 2000). More specifically, probit models allow modelling of certain individual threshold values that, if surpassed, lead to the adoption decision. Examples of such threshold values that were explored in the literature included firm size and purchasing costs.

Differences between these modelling techniques are important, as some have a better track record in accurately forecasting diffusion than others (Meade and Islam, 2006). In comparative studies of different forecasting models, the Bass model consistently performed well although other models such as the local logistic, the Gompertz and the logistic model also performed well (Meade and Islam, 2006). I chose the Bass model because it had the greatest number of applications and validation studies (Meade and Islam, 2006). Other than that, the choice of the Bass model over other well-performing models was, to a certain extent, arbitrary. It is therefore important to observe some quality-control. Criteria for good practice that should be observed in making predictions about product growth curves were described as (Meade, 1984):

- Model validity: there should be an upper bound to the saturation level. This could naturally be set at 100% of the eligible population, but is likely to be lower than that in most settings, given that alternative technologies are available. If the saturation level is not carefully estimated, this could result in a large bias in the resulting diffusion curve.
- Statistical validity: estimated model parameters should be tested for significance. Where model parameters are not obtained from observed data but using prediction, Bayesian methods can be used as an alternative to significance testing. This could entail the use of Bayesian reference priors, for example elicited from expert opinion. This is important because it increases the chance that true uptake values are captured in the distribution.
- Demonstrable forecasting ability and validity: contextual plausibility should be demonstrated and some measure of uncertainty shown. Again, this can help ensure that true uptake values are reflected in the estimates. Contextual plausibility could be established by eliciting the opinions of experts, that is, of people who may be involved in the purchase of the technology.

3.4.2 Bass model extensions useful for predicting diffusion prior to technology introduction

Since Bass' model in 1969, many adaptations to this model have been explored. Meade and Islam (2006) provide a useful overview of the different extensions and adaptations. These encompass extensions to incorporate diffusion in different countries, different technology generations and explanatory variables. In this thesis I focus on England, and also on the first generation device of EIS only, making the first two mentioned extensions irrelevant. The inclusion of explanatory variables, however, may be of interest.

One of the objectives of this thesis was to quantify the effect of implementation strategies on diffusion. This would require the elicitation of different diffusion curves, one for a scenario in which no implementation strategies are used, and multiple other diffusion curves for the number (and combinations) of different implementation strategies. That means, for each scenario, Bass model parameters *p*, *q* and *m* would need to be estimated. Including implementation strategies as explanatory variables in the Bass model could potentially help in reducing the number of quantities to elicit. It would therefore be desirable to establish the effects of different implementation strategies and different scales of these implementation strategies on Bass model parameters *p*, *q* and *m*. There is a small literature on explanatory variables impacting on diffusion that was summarised by Meade and Islam (2006). While a few studies attempted to make Bass model parameters *p*, *q* and *m* functions of explanatory variables such as advertising and price (those were the two main factors included as explanatory variables), Bass, Krishnan & Jain (1994) proposed the introduction of a multiplicative term, summarising the effect of explanatory variables in what they called the 'generalised Bass model' (GBM). The authors demonstrated that adding a multiplicative term comprising both price and advertising strategies performed better in several retrospective case studies than estimating the Bass model parameters as functions of price and advertising separately. The advantage of the GBM was seen in that constant environmental effects led to the GBM reducing to the 'basic' Bass model. However, in the GBM, external strategies only affect parameters *p* and *q*, but the total attainable number of adoptions *m* remains unaffected. Assuming that some implementation strategies make a technology more attractive for potential purchasers, this is an unrealistic feature of the GBM.

Subsequent studies found that, rather than predicting explanatory variables, incorporating the time-varying nature of all Bass model parameters could improve forecasting accuracy (Putsis, 1998); and that there was little evidence of an improvement in

forecasting accuracy because of the inclusion of explanatory variables (Bottomley and Fildes, 1998). The authors thought that this may be partly owed to the fact that any changes in those explanatory variables were predictions at the time of product forecasting.

While it is possible to incorporate a functional relationship of different explanatory variables and diffusion in the Bass model, the limitations are that a) it is not clear whether this would improve forecasting accuracy compared to the elicitation of different scenarios; and b) it does not appear to reduce the number of quantities to elicit. These developments would be more complex and for the purposes of this study not necessary.

3.4.3 Addressing limitations of the Bass model

There are problems associated with the use of the Bass model when used to predict diffusion with no data available. One is that estimates of the Bass model parameters are correlated with each other so that estimating them from univariately elicited summaries may yield unrealistic diffusion curves. This will be explained in more detail in Chapter 6. Another problem is the discrete nature of uptake data which is represented in a time-continuous function in the classic Bass model (Equation 3.1). This could lead to over- and under-estimation of the difference equation from one period to another, given certain values for parameters p and q. While this is not a problem in the majority of cases, it could result in an 'oscillating' cumulative diffusion curve in a small number of cases where parameters p and q adopt a combination of extreme values. This will also be explained in more detail in Chapter 6.

In search of a solution to this problem, I found a paper by Satoh (2001) that addressed these problems. This did not come up in the previous search because that search was specifically designed to find publications on pre-launch forecasting. I therefore performed a google scholar search on this issue and identified the paper by Satoh (2001). The Satoh approach is a transformation to the Bass model that applies a discrete analogue of a quadratic first order equation proposed by Hirota (1979) to the Bass model to make it discrete in time, rather than continuous in time. This is supposed to avoid the problem of 'oscillating' diffusion curves in extreme values for p and q. The Satoh approach is described in more detail in Chapter 4 and is applied and tested for its performance in terms of predicting diffusion in comparison with the classic Bass model in Chapter 6.

3.4.4 Parameter estimation

Although I decided against the use of diffusion data from analogous technologies or from meta-analysis to inform the diffusion curve of EIS, it is worth noting sources of evidence on the diffusion model in different technologies, for purposes of comparison with the elicited diffusion curves and future reference when analysts lack the time for an elicitation exercise.

A meta-analysis of a variety of diffusion curves from the 1950s to 1980s was conducted by Sultan et al. (1990). Mean values for p and q were reported, however, significant variation was observed between technologies and no standard deviations were reported. The authors explained that part of the variation (up to 50%) was systematic and explainable. Results included that the main factors influencing parameter estimates were the type of innovation, the country and the variables included. Sultan et al. (1990) specifically highlighted that diffusion was not an 'automatic' process and that inclusion of environmental variables such as marketing activities was clearly indicated. The authors also reported ranges for p and q that had been observed in other studies: p ranged from 0.000021 to 0.03297 and q from 0.2013 to 1.67260. These may give an indication of what value ranges p and q will probably fall into but also highlight the great variation that was observed with different types of technologies. The results from their own meta-analysis were mean values of $p = 0.03$ and $q = 0.38$. In another meta-analysis of 52 consumer durables in 28 countries (van den Bulte and Stremersch, 2004), the authors found mean values of $p = 0.027$ and $q = 0.419$.

In health technologies, reported Bass model parameters are hard to find but there are a few applications in health that are potentially relevant. In a study by Sillup (1992), the Bass model was used to retrospectively predict product growth of several medical technologies, including computerised tomography scans, MRI and ultrasound. Unfortunately, relevant parameter values were not reported. Another application of the Bass model was in a new neuro-monitoring device that reported the concentration of certain biomarkers in the blood (Gobok et al., 2009). At the development stage, no decision on the use of this technology had been made and a variety of settings were evaluated using the Bass model. The settings included stroke monitoring in intensive care units, traumatic brain injury monitoring, sepsis monitoring and neonatal paediatric monitoring. The authors used estimates for parameters p and q for ultrasound imaging in the medical industry (which were reported to be $p = 0.0013$ and $q = 0.6196$ (Lilien and Rangaswamy, 2004) based on sales data available from 1965 to 1978) and adapted these using a scoring system to reflect the potential differences in diffusion of the neuromonitoring device. They then arrived at their final predicted values for sepsis, stroke, neonatal monitoring and traumatic brain injury, respectively: $p = 0.0052$ and

$q = 0.8328$, $p = 0.0041$ and $q = 0.7062$, $p = 0.0009$ and $q = 0.6995$, $p = 0.0051$ and $q = 0.8061$. *P* and *q* values for another health technology were reported by Meade and Islam (2006) who showed predicted time-varying *p* and *q* values for the adoption of mammography in US hospitals.

3.5 Summary

This chapter identified no feasible and existing methods for obtaining probabilistic estimates of health technology diffusion prior to technology introduction, when no uptake data were available. Synthesising available evidence from the UK health technology sector would have been associated with challenges difficult to overcome, because of heterogeneity of the data, small numbers of analogous technologies and short lengths of time series. Forecasting methods have mainly focussed on guessing by analogy, a method that, given the limited available data in the health technology sector, was not deemed suitable. While expert judgements were another method occasionally used for pre-launch forecasts, the resulting diffusion estimates reported in the literature were not probabilistic and issues were reported in converting elicited quantities into parameters of the most established diffusion model, the Bass model of new product growth. It was highlighted that formal diffusion models such as the Bass model should provide the structure for an expert judgement exercise. The Bass model was chosen for further research because it was well established in forecasting tasks, values of model parameters were available for many different technologies including health technologies, and its fit with real data was demonstrated in many publications.

The reviews presented in this chapter were important in identifying a research gap and developing the scope and methods for this thesis. The consulted diffusion model and forecasting literature together with a review of available UK health technology diffusion data suggested that replicating existing methods was not feasible or would likely not yield accurate estimates. Instead, this chapter established the need for an expert judgement method that can provide probabilistic estimates of diffusion and that can elicit quantities that can be transformed to yield the Bass model parameters.

A strength of this chapter was that both the use of synthesised diffusion data and methods for estimating individual technology diffusion were considered. I was therefore confident that the decision to develop an expert elicitation study using a diffusion model was worth further research efforts, providing scope for improving forecasting accuracy in health

technologies. A limitation is that it was not entirely clear whether the Bass model was the best diffusion model for such an exercise. However, its advantages were clearer than those associated with other diffusion models and there was indication that its limitations could potentially be addressed using the Satoh adaptation.

In conclusion, the reviews presented in this chapter informed the decision to focus on developing a formal elicitation of expert opinion method to obtain probabilistic diffusion estimates based on the Bass model for individual health technologies prior to technology introduction.

CHAPTER 4. MODELLING FRAMEWORK FOR INCORPORATING DIFFUSION ESTIMATES IN HEALTH ECONOMIC METHODS

4.1 Background

In Chapter 2, four literature reviews were conducted through which I identified how health technology diffusion could affect health economic outcomes. While other effects are possible, the main areas to be studied in this thesis were the value of implementation and the effect of diffusion on health technology pricing through experience curves. In Chapter 3, I identified an established diffusion model: the Bass model of new product growth.

The aim of this chapter was to develop a model framework that allows incorporation of diffusion estimates in health technology assessments. For this, I develop several extensions to the expected value of information and implementation (EVII) model in Section 4.2. In Section 4.3, I present the dynamic cost-effectiveness analysis (DCEA) and experience curve model. In Section 4.4, I present the diffusion models used for this thesis: the Bass model of new product growth and the Satoh adaptation of it. I end with a discussion and conclusion in Section 4.5. The place of this chapter within this thesis is illustrated in Figure 1.2 in Chapter 1.

4.2 The dynamic value of information and implementation framework

As was identified in Chapter 2, there is an existing framework, the unified EVII developed by Fenwick et al. (2008) that analyses the expected opportunity loss associated with low implementation and uncertainty. As was mentioned before, the authors developed a four-state model where both implementation and information could adopt the states ‘current’ and ‘perfect’. The limitations of this existing framework were (a) that it did not allow for states in between current and perfect; (b) that the potential effect of evidence generation schemes on implementation was ignored; and (c) that the framework did not allow for the dynamics of diffusion. Existing extensions identified in the review in Chapter 2 did not fully address these issues. One study attempted to relax assumption (a) but did not do so correctly (Andronis and Barton, 2016). Another framework was identified that relaxed assumption (b) (Willan and Eckermann, 2010). While this was done successfully, the Willan and

Eckermann work had the limitation that the effect of research on implementation was modelled using strong assumptions (a threshold value of the z-static at which implementation would jump to a certain value), and that implementation was static. Three studies attempted to relax assumption (c) but did not do so using established diffusion models (Andronis and Barton, 2016, Whyte et al., 2016, Faria et al., 2014).

The aim for this section was to develop the unified EVII framework further to enable the analysis of particular implementation strategies: first, by capturing their effect on implementation and second, to assess their value in terms of the reduction of the value of implementation. In this framework, I also allow research studies to have an effect on implementation. A further objective was to incorporate the dynamics of health technology diffusion in this framework and to allow for implementation measures to take effect with delayed timings. To achieve this, I illustrate the dynamic value of implementation framework in a hypothetical example.

4.2.1 The existing EVII framework

4.2.1.1 The expected value of implementation

The *Expected Value of Perfect Implementation (EVPIM)* was developed by Fenwick et al. (2008). This framework described the expected opportunity loss associated with low implementation of a cost-effective health technology. The authors also described the relationship between further evidence generation and investment in implementation measures, treating them as two alternative strategies with the potential to add value to a health system. The framework uniquely enables a reimbursement authority to assess where the greatest potential for adding further value lies: in improving implementation or in reducing uncertainty through further generation of evidence.

As is shown in Equation (4.1), the EVPIM consists of two components. The right hand side after the minus sign is the uptake-weighted mean of the expected net benefits of all technologies in one appraisal. This mean expected net benefit that is attainable for the health system under current (or anticipated) uptake levels is then subtracted from the maximum expected net benefit among the different decision options in the appraisal. It is worthwhile mentioning here that when I mention the net benefit in this thesis I generally refer to the net monetary benefit of a technology, although all of the following value measures could also be reflected in terms of net health benefits.

$$EVPIM = max_d \mathbb{E}_\theta NB(d,\theta) - \sum_{d=1}^{D} \rho_d^C \mathbb{E}_\theta NB(d,\theta) \quad (4.1)$$

where ρ_d^C is the probability of implementing technology *d* with current information, $NB(d,\theta)$ is the expected net monetary benefit of technology *d* given the uncertain model input parameters θ.

The EVPIM is important because, if unresolved (that is, when full implementation is not achieved), it gives a value to inefficiencies in the health system that are caused by cost-ineffective technologies remaining in use.

4.2.1.2 The expected value of information

The expected value of information, or the payer uncertainty burden as it was most recently called (Grimm et al., 2016), is an established concept that describes the expected opportunity loss associated with a decision made under uncertainty. *Expected Value of Information (EVI)* methods estimate the magnitude of uncertainty on a cost or health scale and can help identify the drivers of that uncertainty, thus determining the areas in which further research may be required (Briggs et al., 2006). The *Expected Value of Perfect Information (EVPI)* is the expected value of reducing all uncertainty present in a decision (Briggs et al., 2006). The EVPI is calculated by subtracting the maximum expected net benefit among the different decision options in the appraisal from the expectation of the maximum net benefit in each iteration of the probabilistic sensitivity analysis (PSA). Note that the part of the equation after the minus sign in Equation (4.2) equals the part of the equation before the minus sign in Equation (4.1).

$$EVPI = \mathbb{E}_\theta max_d NB(d,\theta) - max_d \mathbb{E}_\theta NB(d,\theta) \quad (4.2)$$

The EVPI is relevant because most technology appraisals are associated with considerable uncertainty that may result from an evidence base that has not been fully developed. Uncertainty may also stem from other sources of evidence such as the value assigned to certain health states or the cost of a service outside clinical trials. When a decision is made under uncertainty, there is a risk of making the 'wrong' decision, that is, recommending a health technology that does not provide the true maximum net benefit

(Grimm et al., 2016). The cost of making this decision under uncertainty, the EVPI, is therefore composed of the risk of making the 'wrong' decision and the consequences of making that decision in terms of costs or health foregone (Grimm et al., 2016).

4.2.1.3 The expected value of perfection

When there is uncertainty and low implementation, the payer faces a combined burden of uncertainty and implementation that is described by the *Expected Value of Perfection (EVP)* (Fenwick et al., 2008). The EVP, being the sum of the EVPI and the EVPIM, is calculated by subtracting the second part of Equation (4.1) from the first part of Equation (4.2), which is shown in Equation (4.3).

$$EVP = \mathbb{E}_\theta max_d NB(d,\theta) - \sum_{d=1}^{D} \rho_d^C \mathbb{E}_\theta NB(d,\theta) \qquad (4.3)$$

The EVP is important because it highlights the opportunity loss experienced by a health system if a decision is made without resolving uncertainty and without improving implementation. The goal of a decision-maker could be to recommend measures to resolve the EVP, or reduce it as far as possible.

4.2.2 Extension to assess the value of specific implementation measures

I propose an extension to the Fenwick et al. (2008) framework to assess the *expected value of specific implementation measures (EVSIM)* by allowing for states other than current and perfect. The term EVSIM was proposed by Fenwick et al. (2008) but the authors did not demonstrate how this was calculated.

States other than current or perfect have been incorporated in value of information methods for a long time by calculating the *expected value of sample information (EVSI)*, in which particular designs of research studies can be assessed. A parallel can be drawn between assessing the value of sample information and the value of implementation measures. For this reason, I first discuss how the EVSI is commonly calculated and will then present the calculation of the EVSIM.

To calculate the EVSI, it is required to simulate the data to be collected in the future. To do this, one needs to specify a statistical model that describes the data to be collected. This

model can be used to draw from in a large number of simulations, to generate the data to be collected conditional on the PSA results of the parameter(s) of interest. To add the simulated data to the evidence available so far, it is necessary to calculate the posterior distribution for the parameter(s) of interest (Brennan and Kharroubi, 2007, Ades et al., 2004, Strong et al., 2015). This can be a complex exercise because Markov Chain Monte Carlo simulation was typically required to calculate the posterior when an analytic solution was not available (Brennan and Kharroubi, 2007). Recent developments have simplified the step of calculating the posterior by using nonparametric regression methods (Strong et al., 2015). The EVSI is then the difference between the expected value of the best decision with sample information and the best decision's expected value without sample information, as shown in Equation (4.4):

$$EVSI = \mathbb{E}_X\left[max_d \mathbb{E}_{\theta|X}\{NB(d|\theta)\}\right] - max_d[\mathbb{E}_\theta\{NB(d,\theta)\}] \quad (4.4)$$

where data X are informative for the input parameters θ.

Prior to calculating the EVSI, it is therefore useful to determine how much individual or grouped input parameters used in the cost-effectiveness model contribute to decision uncertainty. *Expected Value of Perfect Parameter Information (EVPPI)* calculations can be performed to achieve this. The EVPPI indicates the expected value of reducing all uncertainty on a specified parameter or group of parameters (Strong et al., 2014) and is presented in Equation (4.5):

$$EVPPI = \mathbb{E}_{\theta_i}\left[max_d \mathbb{E}_{\theta_{-i}|\theta_i}\{NB(d,\theta_i|\theta_{-i})\}\right] - max_d \mathbb{E}_\theta NB(d,\theta) \quad (4.5)$$

where θ_i is the parameter of interest and θ_{-i} are the remaining input parameters.

Assessing the value of specific implementation measures is more straightforward. Instead of a simulation of data, all that is needed are estimates of the implementation levels achievable before and after implementation. The EVSIM is then the difference between the expected net benefits weighted by the achieved implementation before and after implementation measures, as shown in Equation (4.6):

$$EVSIM = \sum_{d=1}^{D} \rho_d^{IM} \mathbb{E}_\theta NB(d,\theta) - \sum_{d=1}^{D} \rho_d^{C} \mathbb{E}_\theta NB(d,\theta) \quad (4.6)$$

where ρ_d^{IM} is the with-implementation measure implementation of technology *d*.

If implementation had effects on the net benefit, for example in the case of vaccinations or where learning effects could improve effectiveness, this could be incorporated into Equation (4.6) by making the net benefit function conditional on implementation.

Uncertainty over uptake is not considered in this equation, but can be incorporated by simulating the EVSIM drawing from the probability distributions over ρ_d^{IM} and ρ_d^C. When this is done, it is possible to perform a value of information analysis on the implementation estimate using the EVSIM function as the objective function and performing an EVPPI analysis on the implementation parameter.

The EVSIM would need comparing with costs of the implementation strategy. It is then important to consider not only fixed costs, but also potential variable costs that arise with the uptake. In such a case, the cost function would also contain an estimate of ρ_d^{IM}.

4.2.3 Disentangling the effect of research on value of implementation and the effect of implementation on value of information

The assumption made by Fenwick et al. (2008) and subsequent papers identified in Chapter 2 (with a few exceptions mentioned in the following) was that the effect of research studies on the EVPIM was zero. The authors acknowledged that research could have an effect on implementation and that it would be worthwhile exploring the value of implementation of research studies. They also provided an exploratory analysis in which they let information affect implementation, which I describe below. Only two studies have incorporated the effects of information on implementation at the time of writing. The study by Willan and Eckermann (2010) focused on establishing a functional relationship between the strength of research evidence and potential implementation. The recent study by Andronis and Barton (2016) developed the implementation-adjusted EVSI but interpreted it incorrectly. I will demonstrate this in this section, by investigating the effect of research on the value of implementation.

To fully understand the effects of research on implementation, it is useful to examine in more detail how the EVPI and EVSI work. In EVPI calculations, we follow two simple steps: 1. we assess the value of a decision we would make with full information. In fact, given full information, we would choose the optimal decision option every time (for example, in every run of the PSA, and given that we are rational and utility maximising). 2. We compare the expected net benefit of making this optimal decision every time with the expected net

benefit of choosing one decision option over the others once, based on current information expected values, and then stick with this decision every time (that is, in every run of the PSA). It follows that the value of information is entirely derived from the expected value of the discrepancy between the optimal and sub-optimal decision (the incremental net benefit) and the number of times a sub-optimal decision would be chosen under current knowledge.

Incorporating implementation in an EVPI formula alters this and makes the number of times we make the sub-optimal decision independent of the expected net benefit. Implementation therefore removes the value of the reduction of uncertainty component. So, if we incorporate technology implementation in the EVPI equation, we are not capturing the effect of uncertainty, because now, despite having full information on which technology to adopt, we are not choosing this technology every time. Equally, despite believing in the maximum expected net benefit of one technology over the other under current knowledge, we do not choose this technology every time but instead only choose it some of the time. The term 'realisable EVPI' that was employed by Fenwick et al. (2008) is therefore misleading as it may evoke the impression that the concept describes the value of information of a research study under imperfect implementation rather than the value that a research study contributes through its effect on implementation. The authors' explanation of the realisable EVPI as *'the expected value of research that is realizable without actively undertaking strategies to change implementation'* is perhaps more intuitive than the term 'realisable EVPI'. Equally, and probably following from a misunderstanding of the term, the Andronis and Barton (2016) interpretation of the implementation-adjusted EVSI is incorrect.

This is important: it means that by including implementation in EVPI, EVPPI or EVSI calculations we are not assessing the value of reducing uncertainty any longer. We instead capture the value of the reduction of uncertainty in terms of improving implementation. Both effects are dissimilar and unrelated. I therefore suggest separating these two effects by making a clear distinction between value of information and implementation measures. To make this distinction, when research can affect implementation, the expected value of perfection which previously consisted of the EVPI and EVPIM can now be divided up into three components: 1. The expected value of perfect information. 2. The expected value of perfect implementation resolvable by research (I call this the Research EVPIM, or shorter, the $EVPIM^R$). 3. The expected value of perfect implementation resolvable through other implementation measures ($EVPIM^{IM}$).

The EVP, being a measure of both values of information and implementation, can be reduced by implementation measures and research. When perfect implementation of the most cost-effective technology has been achieved without doing any research, the EVP

reduces to the EVPI. (Of course, the EVP with perfect implementation, or the EVPI, may be reduced if part of the activities to achieve perfect implementation was research.) Mathematically, this is because the second part of Equation (4.3) (shown on the left hand side below) would become the second part of Equation (4.2) (shown on the right hand side below), thus converting Equation (4.3) (the EVP) into Equation (4.2) (the EVPI):

$$\sum_{d=1}^{D} \rho_d^C \mathbb{E}_\theta NB(d,\theta) = 1 * \mathbb{E}_\theta NB(d^*,\theta) + \sum_{d'}^{D'} 0 * \mathbb{E}_\theta NB(d',\theta) = max_d \mathbb{E}_\theta NB(d,\theta),$$

where d^* is the decision option expected to be most cost-effective, d' is a set of other cost-ineffective decision options and $\rho_d^C = 1\ for\ d^*, \rho_d^C = 0\ for\ d'$.

When perfect information has been achieved first, in contrast, the EVP in theory reduces to the EVPIM, because all uncertainty is resolved and all that is left to do is to improve implementation. Mathematically, this can be shown by the first part of Equation (4.3) (EVP, shown on the left hand side below) that would become the first part of Equation (4.1) (EVPIM, shown on the right hand side below):

$$\mathbb{E}_\theta max_d NB(d,\theta) = max_d NB(d,\theta) = max_d \mathbb{E}_\theta NB(d,\theta).$$

However, effects of research on implementation may result in a with-research residual EVP that is lower than the original EVPIM. This is because a part of the EVPIM, the *Research EVPIM*, can be resolved through research. This Research EVPIM is what was called the realisable EVPI by Fenwick et al. (2008). The distinction may be subtle, but it is important: the term realisable EVPI implies that the full value of the EVPI cannot be reached due to the low implementation. This is, however, not the case, as the Research EVPIM could be larger than the EVPI. This was shown in the exploratory analysis by Fenwick et al. (2008) and is also demonstrated in the following equations. The Research EVPIM is shown in Equation (4.7):

$$EVPIM^R = \mathbb{E}_\theta \sum_{d=1}^{D} \rho_d^P NB(d,\theta) - \sum_{d=1}^{D} \rho_d^C \mathbb{E}_\theta NB(d,\theta) \qquad (4.7)$$

The fact that the Research EVPIM can exceed the expected value of perfect information (and indeed exceeded it in the exploratory analysis by Fenwick et al. (2008)) is explained by the following: the first term of the Research EVPIM Equation (4.7) ($\mathbb{E}_\theta \sum_{d=1}^{D} \rho_d^P NB(d,\theta)$) converges to the first term of the EVPIM Equation (4.1) ($max_d \mathbb{E}_\theta NB(d,\theta)$) when implementation of the cost-effective technology approaches a level of 100%; and the second term of both Research EVPIM and EVPIM equations are equal. The value of research can therefore be much larger than the EVPI alone and the term realisable EVPI is therefore

misleading. For this reason, I continue to use the expression 'expected value of perfect implementation that is resolvable with research' or short the Research EVPIM, and the contrasting 'expected value of perfect implementation that is resolvable through other implementation measures' or the 'implementation measure (IM) EVPIM'. The EVPIM is the sum of the Research EVPIM and the implementation measure EVPIM:

$$EVPIM = EVPIM^{R} + EVPIM^{IM} \tag{4.8}$$

Fenwick et al. (2008) further established that when implementation with perfect information equals implementation under current information ($\rho_d^P = \rho_d^C$), the Research EVPIM equals zero. This is because the expected value of a decision under perfect information is the same as under imperfect information (Fenwick et al., 2008). When implementation at perfect information ρ^P equals one, the Research EVPIM equals the EVPIM. Of course, this holds only as long as both measures use the same value of current implementation ρ_d^C as the baseline.

4.2.4 Extension to calculate the expected value of research

It follows from the above that a research study can be assessed in terms of both its effect on implementation and its effect on information. Willan and Eckermann (2010) implicitly calculated the joint effect of both in their EVSI analysis but did not make the different effects of implementation and information explicit. Modelling these different effects explicitly, however, results in greater transparency that could be useful for decision-makers. I therefore propose the E*xpected Value of Research (EVR)*, which measures the effect of research on both the reduction of uncertainty and the change in implementation. The EVR is simply the sum of the EVSIM of the research study and the EVSI of the same study, shown in Equation (4.9):

$$EVR = EVSIM^{R} + EVSI \tag{4.9}$$

The EVR is thus:

$$\begin{aligned} EVR = \sum_{d=1}^{D} \mathbb{E}_X[\rho_d^X \mathbb{E}_\theta NB(d,\theta)] - \sum_{d=1}^{D} \rho_d^C \mathbb{E}_\theta NB(d,\theta) \\ + \mathbb{E}_X\left[max_d \mathbb{E}_{\theta|X}\{NB(d|\theta)\}\right] \\ - max_d[\mathbb{E}_\theta\{NB(d,\theta)\}] \end{aligned} \tag{4.10}$$

where ρ_d^X is implementation of technology d with availability of data X.

The Research EVSIM can be written similarly to any other EVSIM because the effects of the reduction of uncertainty cancel out with the inclusion of implementation, as is shown in Equation (4.11). The difference to EVSIMs associated with implementation strategies, however, is that diffusion is also influenced by the outcomes of the study, which are not pre-planned. ρ_d^X is therefore associated with uncertainty and a dependent covariate of net benefit.

$$\begin{aligned} EVSIM^R = \mathbb{E}_X \left[\sum_{d=1}^{D} \rho_d^X \mathbb{E}_{\theta|X}\{NB(d,\theta)\} \right] \\ - \sum_{d=1}^{D} \rho_d^C \, \mathbb{E}_X \left[\mathbb{E}_{\theta|X}\{NB(d,\theta)\} \right] \\ = \left[\sum_{d=1}^{D} \mathbb{E}_X [\rho_d^X \mathbb{E}_{\theta|X}\{NB(d,\theta)\}] \right] \\ - \sum_{d=1}^{D} \rho_d^C \, \mathbb{E}_X \left[\mathbb{E}_{\theta|X}\{NB(d,\theta)\} \right] \\ = \left[\sum_{d=1}^{D} \mathbb{E}_X [\rho_d^X \mathbb{E}_{\theta|X}\{NB(d,\theta)\}] \right] \\ - \sum_{d=1}^{D} \rho_d^C \, \mathbb{E}_\theta NB(d,\theta) \end{aligned} \tag{4.11}$$

If we were only interested in the effect of the strength of research evidence on diffusion, assuming that study results were more or less the same as our prior expectation of them, then ρ_d^X can be treated as independent of the net benefit function. Equation (4.11) then reduces to:

$$
\begin{aligned}
EVSIM^R = \; & \mathbb{E}_X \left[\sum_{d=1}^{D} \rho_d^X \mathbb{E}_{\theta|X}\{NB(d,\theta)\} \right] \\
& - \sum_{d=1}^{D} \rho_d^C \, \mathbb{E}_X \left[\mathbb{E}_{\theta|X}\{NB(d,\theta)\} \right] \\
& = \left[\sum_{d=1}^{D} \rho_d^X \mathbb{E}_X [\mathbb{E}_{\theta|X}\{NB(d,\theta)\}] \right] \\
& - \sum_{d=1}^{D} \rho_d^C \, \mathbb{E}_X \left[\mathbb{E}_{\theta|X}\{NB(d,\theta)\} \right] \\
& = \left[\sum_{d=1}^{D} \rho_d^X \mathbb{E}_\theta \{NB(d,\theta)\} \right] - \sum_{d=1}^{D} \rho_d^C \, \mathbb{E}_\theta NB(d,\theta)
\end{aligned}
\tag{4.12}
$$

More intuitively, and as was discussed in the previous section, the effect of uncertainty reduction is now not captured in the Research EVSIM: including implementation has removed the term that accounted for the effect on uncertainty, which is the difference between the maximisation and the number of times the sub-optimal decision would have been made based on choosing the decision option with the highest expected net benefit under current information. With the EVSI, the maximum expected net benefit conditional on the new data was chosen. With the EVSIM of research, however, the expected net benefit is accrued for a certain part of the population, and the newly generated data plays no role in its calculation, except in the change of the implementation estimate. This clarifies where Andronis and Barton (2016) went wrong: what they called the implementation-adjusted EVSI was in fact the Research EVSIM shown in Equation (4.11). Their interpretation of it was that the effect of research on the EVPI would be dampened by low implementation. In contrast, I have shown above that the Research EVSIM has to be understood as the effect of research on implementation. The authors further stated that the implementation-adjusted EVSI would always be lower than the EVSI and even claimed that the value of research was systematically over-estimated if not adjusted by implementation. I show the opposite: if research has an effect on implementation, then its value may be under-estimated when only the EVSI is calculated.

Similar to the EVPPI, the value of resolving all uncertainty caused by one or more parameters can also be assessed in terms of the effect on implementation using the *Research EVSIM for perfect parameter information (*$EVSIM^{Rpp}$) shown in Equation (4.13). This

enables calculating the ceiling value of the potential value of implementation of research studies on certain groups of parameters. As such, its calculation is no different from other EVSIM calculations. The $EVSIM^{Rpp}$ thus answers the question of how much the value of implementation would reduce if we removed all uncertainty surrounding a (group of) parameter(s).

$$EVSIM^{Rpp} = \sum_{d=1}^{D} \rho_d^{Rpp} \mathbb{E}_{\theta_i}\{NB(d,\theta_i)\} - \sum_{d=1}^{D} \rho_d^{C}\, \mathbb{E}_{\theta} NB(d,\theta) \tag{4.13}$$

where ρ_d^{Rpp} is the with perfect parameter information implementation of technology d.

4.2.5 Extension to make the EVII dynamic

The analyses described above took neither implementation dynamics nor timings of research and implementation measures into account. In this section, I describe in what way implementation dynamics can be used in the calculation of the expected value of implementation measures. Calculations of the EVPI, the EVPPI and the EVSI remain unaffected as implementation is not considered in them.

With implementation varying over periods of time, all the value of implementation measures will accrue different values for each period up to the defined decision relevance horizon. These can be shown individually in a tabular overview, such as in the example shown in the first column of Table 4.1. Using the $EVPIM^{R}$ as an example, if utilisation of intervention A was at 2% in the first year, then utilisation of the only other two comparators B and C in the same model would be at 49% each (assuming equal utilisation for the two remaining technologies, due to a lack of knowledge that suggests otherwise), and the $EVPIM^{R}$ would be calculated accordingly. In the second year, implementation of A might be at 10%, resulting in 45% of implementation of B and C, and so on.

To obtain these different values of the $EVPIM^{R}$ in each period of time, I let the implementation estimate be time variable as shown in Equation (4.14). The same principles apply to the calculation of the $EVPIM^{IM}$, the $EVSIM^{Rpp}$ and the EVSIM.

$$EVPIM^{R}{}_{t} = \mathbb{E}_{\theta} \sum_{d=1}^{D} \rho_{t,d}^{P} NB(d,\theta) - \sum_{d=1}^{D} \rho_{t,d}^{C} \mathbb{E}_{\theta} NB(d,\theta) \tag{4.14}$$

where $t=1,\ldots,T$ is the time period up to the defined decision relevance horizon T, in years.

4.2.6 Accruing value measures over population and decision relevance horizon

EVII measures are typically presented accrued over the affected patient population and the decision relevance horizon, for example until there is a new technology class replacing the current one. This is commonly done by simply adding up the values of the EVPI (or any other value measure) over the future time periods up to the decision relevance horizon. Using the example of the EVPIM, it can be multiplied by the discounted population in each period as is shown in Table 4.1.

Table 4.1 Tabular overview of example dynamic EVPIM

Time *t*	Individual (per person) $EVPIM$	Discounted population affected	$EVPIM$ accrued over population
1	$EVPIM_1$	$\delta_1\pi_1$	$\delta_1\pi_1 EVPIM_1$
2	$EVPIM_2$	$\delta_2\pi_2$	$\delta_2\pi_2 EVPIM_2$
3	$EVPIM_3$	$\delta_3\pi_3$	$\delta_3\pi_3 EVPIM_3$

where δ is factor describing discounting at $\frac{1}{(1+r)^t}$, with r being the discount factor and π_t is the number of affected patients in period *t*.

For the evaluation of specific implementation measures, the use of the discounted population measure without any adjustment for implementation would over-estimate the value of implementation. This is because only the patients that will receive the most cost-effective technology will benefit from the success of the implementation measure. When calculating the per period EVSIM, the population estimate should therefore be adjusted by the achievable implementation in each period, as illustrated in Table 4.2. This was not done in Table 4.1 because, arguably, when calculating the EVPIM, the decision-maker is interested in the maximum value of resolving the EVPIM.

Table 4.2 Tabular overview of example dynamic EVSIM – adjusting the patient population by implementation

Time *t*	Individual (per person) EVSIM	Discounted population affected	Implementation of cost-effective technology	EVSIM accrued over implementation-adjusted population
1	$EVSIM_1$	$\delta_1\pi_1$	$\rho_{1,d}$	$\rho_{1,d}\delta_1\pi_1 EVSIM_1$
2	$EVSIM_2$	$\delta_2\pi_2$	$\rho_{2,d}$	$\rho_{2,d}\delta_2\pi_2 EVSIM_2$
3	$EVSIM_3$	$\delta_3\pi_3$	$\rho_{3,d}$	$\rho_{3,d}\delta_3\pi_3 EVSIM_3$

It is furthermore noteworthy that the EVSIM calculation in the future relies on the expected net monetary benefit in the future periods. If net benefit changes in future periods, for example due to the incorporation of experience curves, learning curves or other changes in model parameters over time, then this can easily incorporated in the EVSIM calculation by making net benefit time-dependent.

To facilitate comparison with static per patient analysis, as opposed to above population analysis, which is typically presented for the first year, the average taken over all periods can be used to represent an annual per patient estimate of the dynamic value measures. Using the EVSIM as written in Equation (4.6) as an example, I do not take discounting into account in the calculation of the average future EVSIM because it is designed to compare with the static EVSIM which does not consider discounting:

$$EVSIM^{dyn} = \frac{1}{T}\sum_{t=1}^{T} EVSIM_t \tag{4.15}$$

If the population value of this $EVSIM^{dyn}$ was of interested, it can be calculated by accruing it over the uptake-adjusted population. I assume here that the devised implementation measure is designed to increase implementation for the intervention with the largest expected net benefit. The discounted population is therefore adjusted by the implementation in *t* that is attainable for the technology in question conditional on the implementation measure being used:

$$pop.EVSIM^{dyn} = \frac{1}{T}\sum_{t=1}^{T} \rho^{IM}_{t,d_{\text{NB.max}}}\pi_t\, EVSIM_t \tag{4.16}$$

with $\rho^{IM}_{t,d_{\text{NB.max}}}$ being the attainable implementation (with the implementation measure in question) in each time period *t* of the technology with the maximum expected net benefit.

Population estimates can also be adjusted for implementation in EVPI calculations. This can make sense, for example, in the case that implementation was at a certain level and it was unlikely that anything could be done about it. The realistic EVPI if implementation is a given can only be accrued for the part of the patient population that will receive the cost-effective technology:

$$pop.EVPI^{dyn} = \frac{1}{T}\sum_{t=1}^{T} \rho^{P}_{t,d_{\text{NB.max}}}\, \pi_t\, EVPI_t \tag{4.17}$$

with $\rho^{P}_{t,d_{\text{NB.max}}}$ being the implementation at perfect information of the technology with the maximum expected net benefit in each time period *t*.

4.2.7 The timing of implementation measures and their effects

When assessing the expected value of implementation measures in a dynamic analysis, the timing of the measure itself and its effects are crucial. These two effects are separate. For instance, if an implementation measure is a research study that takes two years to report from the time of decision-making, the existing EVPI will only be reduced at the time that a new decision is made that is based on that new research. It may also be that the improvement of implementation occurs with a delayed effect. Accounting for these two effects is simple: for the periods until the research reports, the EVII with current information is used. The EVP reduces by the value of sample information at the time that research reports (assuming that a new decision is made straight away) and it reduces further at the time when implementation improvements take effect.

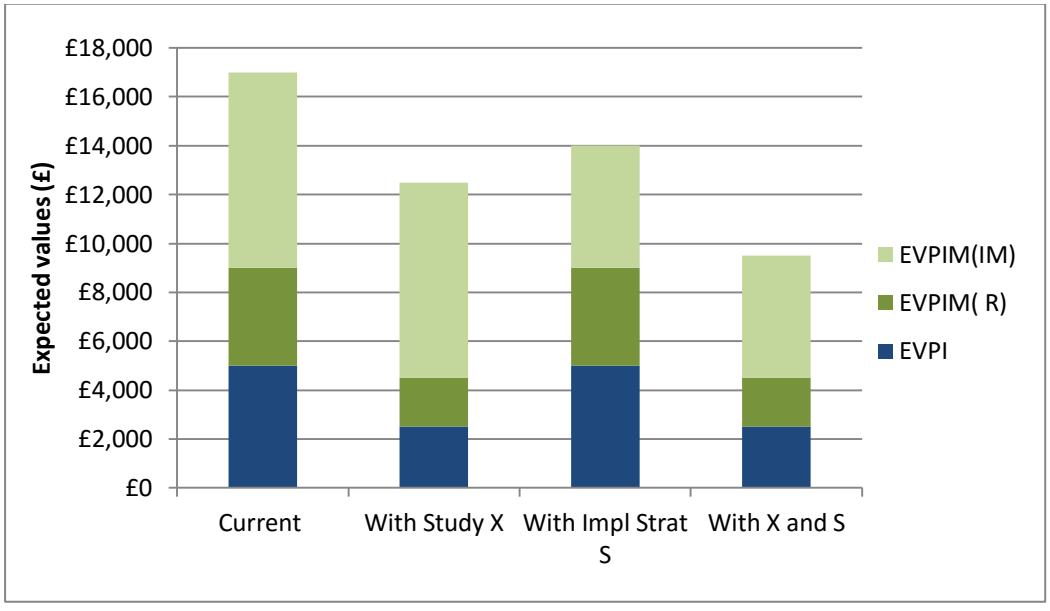
4.2.8 Proposed presentation of the EVII framework using an illustrative example

To illustrate the potential use of the dynamic EVII framework in health technology assessments, I present a hypothetical analysis of an illustrative example technology A. New technology A is cost-effective compared to existing technology B; however, there is uncertainty associated with the decision that is reflected in a positive EVPI. Available

estimates of natural diffusion for technology A suggest that A will not achieve perfect implementation over a decision relevance horizon of 10 years. EVPPI analysis suggests that further research evidence on one parameter θ_i can reduce decision uncertainty and that this exact research study (called Study X) has a positive effect on implementation. Diffusion estimates for a scenario in which this research has been conducted are available. It would also be possible to improve implementation with an implementation strategy S; and diffusion estimates for this scenario are available.

The resulting value of perfection and value of research can then be presented in the EVII analysis chart, which is shown in Figure 4.1. This shows the EVP, split into the EVPI, the Research EVPIM and the implementation measure EVPIM for current information and implementation in the first stacked bar. The second stacked bar shows the residual EVP with Study X and it can be seen that doing this research study would result in a reduction of both the EVPI and the Research EVPIM. Performing implementation measure S results in a reduction of the implementation measure EVPIM, but does not achieve the same overall reduction of the EVP as Study X. Both Study X and Strategy S reduce the EVP the most. With the effects of both being independent of each other, Study X and Strategy S have an additive effect on the reduction of the EVP.

Figure 4.1 The EVII analysis chart of example technology A



4.2.9 Summary

Summarising Section 4.2, I extended the existing EVII framework and its further developments by a) presenting an approach to calculating the expected value of specific implementation strategies; b) reflecting the effects of the reduction of uncertainty on implementation; and c) making the framework dynamic by incorporating diffusion estimates. I conclude that whether the burden on the health system that results from low implementation is large depends on three factors: 1. how low implementation is going to be compared to what it could be, that is, the implementation discrepancy, 2. the magnitude of the incremental net benefit of the new cost-effective technology over its comparators, 3. the size of the patient population that is affected. Furthermore, the values of information and implementation are separate entities, but overlap occurs where research affects implementation. Then, the value of research will be composed of the value of information and the value of implementation of that research study. All expected value of information and implementation measures are summarised in Table 4.3.

Table 4.3 The expected value of information and implementation measures

	Value measure	Description
Uncertainty	EVPI	The value of resolving all decision uncertainty
	EVPPI	The value of resolving all uncertainty related to individual or grouped model parameters
	EVSI	The value of a proposed research study design in reducing uncertainty
Implementation	EVPIM	The value of achieving perfect implementation (from current 'baseline' level)
	$EVPIM^{R}$	The value of improving implementation by resolving all decision uncertainty
	$EVSIM^{(R+IM)}$	The value of improving implementation with a proposed implementation measure or research study design
Both	EVP	The value of achieving perfect implementation and resolving all decision uncertainty
	EVR	The value of a proposed research study design in improving implementation and reducing uncertainty

4.3 The dynamic cost-effectiveness analysis framework and the experience curve model

4.3.1 The experience curve model

As was explained in Chapter 2, experience curves describe empirical evidence for price declines with increasing uptake of technologies. There is ample evidence for such price declines in a variety of different technologies as well as in health technologies in a study on 20 medical devices by Brown et al. (2007). As was described in Chapter 2, experience curves are mainly explained by a technology's competitive situation but potentially also through economies of scale (Brown et al., 2007). Experience curves therefore relate technology price to uptake. More specifically, it has been observed in a number of studies that prices decline to a percentage of the technology's initial price every time initial production volume doubles (Brown et al., 2007):

$$P_{N_t} = \begin{cases} P_{N_0} \; for \; 0 < N_t < 2N_0 \\ \propto^{\beta} \; P_{N_0} \; for \; N_t \geq 2N_0 \end{cases} \tag{4.18}$$

where N_t is the cumulative uptake or sales volume up to period *t*, with P_{N_t} being the price at N_t, P_{N_0} is the price that was set at initial quantity N_0 which is maintained until $N_t \geq 2N_0$, α is the experience curve parameter or the percentage of the technology's initial price with $0 < \alpha < 1$, and β is the number of times that the initial quantity had doubled, with $\beta = log_2 \left[\frac{N_t}{N_0}\right]$.

Equation (4.18) reflects the fact that prices remain stable until the initial production quantity has doubled for the first time. Price is essentially dependent on technology uptake through β, the number of times that the initial quantity had doubled, rather than on time, requiring estimates of diffusion to inform this model.

4.3.2 The dynamic cost-effectiveness analysis model

To incorporate experience curves in health economic evaluation, the cost calculations in the cost-effectiveness model must be performed over several time periods up to the decision relevance horizon. Costs and benefits of health technologies are typically average values assumed to reflect at least one cohort of patients or an average of future cohorts. The standard measure of assessing a technology's value is the ICER which represents average

incremental costs over average incremental quality-adjusted life-years (QALYs) gained with one technology over another.

$$ICER = \frac{c_i - c_j}{e_i - e_j} \tag{4.19}$$

where c_i, c_j and e_i, e_j are the expected costs and effects of interventions *i* and *j*, with $c, e \geq 0$.

The dynamic cost-effectiveness analysis (DCEA) model incorporates the effects of experience curves and diffusion by modelling future periods up to the decision relevance horizon and using the experience curve and uptake models in the dynamic ICER calculation. I assume that, given a positive reimbursement decision, uptake would grow, and given a negative reimbursement decision, the technology would not be implemented at all. Costs in period *t* are now dependent on price and cumulative uptake up to period *t* through the experience curve model. Contrary to other studies (Hoyle, 2011, Hoyle, 2010, Hoyle and Anderson, 2010), I have refrained from weighting the dynamic ICER by uptake as weighting would lead to a blended ICER that assesses a mix of technologies rather than identifying the most efficient technology based on costs and health effects. The dynamic ICER is shown in Equation (4.20).

$$ICER^{dyn} = \frac{\sum_t^T \Delta c_j(P_{N_t})\delta}{\sum_t^T \Delta e_j(t)\delta} \tag{4.20}$$

where $\Delta c(P_{N_t})$ is the difference in costs between interventions *j*, as a function of price and uptake and $e_j(t)$ are effects in each period of time, both summed up over the number of periods up to decision relevance horizon T and discounted at a discount factor of $\delta = \frac{1}{(1+r)^t}$ with *r* as the discount rate, with $c_j(P), e(t) \geq 0, r \geq 0$.

To my knowledge, a price increase with falling uptake has not been observed. This means that no upwards price effects, which may be caused by falling implementation of technology *j* when uptake of technology *i* rises are accounted for in this model.

Changes in the expected net monetary benefit that now are dependent on uptake and experience curve as well as the decision relevance horizon adopted can then be written as in Equation (4.21).

$$NB = \lambda \sum_{t=1}^{T} e_j(t)\delta - \sum_{t=1}^{T} c_j\left(P_{N_t}\right)\delta \tag{4.21}$$

where NB is the net monetary benefit, and λ is the threshold with $\lambda > 0$.

Of course, incorporating experience curves in the model affects the EVPI through the net monetary benefit calculations. It is likely that adding uncertain parameters on future price change and uptake would increase the decision uncertainty reflected in the probabilistic sensitivity analysis. EVPPI analysis can be used to identify to what extent these parameters contribute to decision uncertainty.

4.3.3 Summary

In this section, I have provided a simple extension to cost-effectiveness analysis that allows modelling future periods and reflecting price changes that are precipitated by diffusion. I conclude that estimates of diffusion and the expected price changes conditional on expected diffusion are required to reflect such future price changes in the DCEA framework.

4.4 The diffusion model

Estimates of diffusion are needed for both frameworks presented in Sections 4.2 and 4.3. In Chapter 3, I identified the Bass model of diffusion to be the most established and validated existing model of diffusion. I briefly presented this model in Chapter 3.3 but provide a more detailed presentation on it in this section. In Chapter 3, I also highlighted potential problems with the Bass model used to predict diffusion with no data available and found a proposed solution by Satoh (2001).

As a side note, what is truly needed to inform the EVII and DCEA frameworks is a market share curve over time for each of the decision options considered. This could be represented by, for example, a multinomial logistic model. Diffusion models focus on only one technology. For the remaining technologies in a market or therapeutic area, an assumption therefore needs to be made on the share of the remaining technologies. Because diffusion models have a greater evidence base, I focus on these. However, further exploration of market share models is considered as an important area for future research.

4.4.1 The Bass model of diffusion

The new product growth model developed by Bass (1969) (often simply referred to as the Bass model) is an adapted logistic model that incorporates the effects of 'innovation' and 'imitation'. Bass distinguished between innovators and imitators which he noted was *'consistent with the characterisation of the behaviour of these groups in the social science literature'* (Bass, 1980). Innovators were considered to be those adopters who adopt the technology without being influenced by peers and imitators were those mainly taking the adoption decision because of their peers. Influence not exerted by peers was called external influence and that exerted by peers internal influence. To name examples, external influence could be direct communication by the manufacturer such as advertising; internal influence could be word-of-mouth. The Bass model typically follows an s-shaped curve for the cumulative number of adoptions and a bell-shaped curve for the per-period number of adoptions. The per-period number of adoptions are represented by Equation (4.22):

$$n(t) = p(m - N_\tau) + \frac{q}{m} N_\tau (m - N_\tau) \tag{4.22}$$

where $n(t)$ is the number of per-period adoptions in period *t* and N_τ is the cumulative number of adoptions up to time $\tau = t - 1$, *p* the coefficient of external influence and *q* the coefficient of internal influence, m the total number of attainable adoptions with $m > 0$.

It can alternatively be written as the probability of adoption in each period t:

$$P(t) = p + \frac{q}{m} * N_\tau \tag{4.23}$$

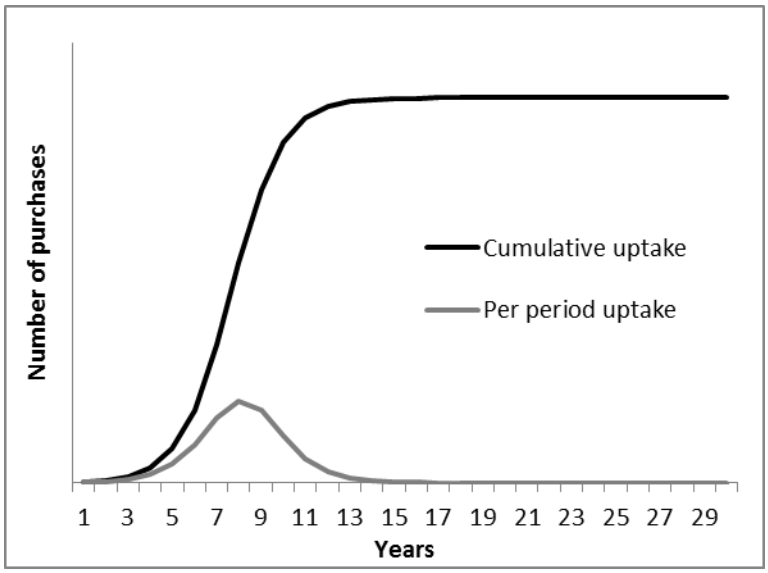
where $P(t)$ is the probability of adoption in period *t*.

It is unclear from the literature what the restrictions for parameters *p* and *q* are. A few meta-analyses reported the ranges that they had found for parameters *p* and *q* (van den Bulte and Stremersch, 2004, Sultan et al., 1990), which were reported in Chapter 3. When I attempted to find natural boundaries of parameters *p* and *q*, I found that the model performed well with parameter values of $0.0001 < p < 0.1$ and $0 < q < 0.9999$. I observed that values outside these ranges may cause the diffusion curve to oscillate. However, this only happens for certain combinations of the *p* and *q* parameters and the impact of this is explored further in Chapter 6.

In a review of diffusion models, Meade and Islam (2006) stated that $p + q$ controls the scale and that $\frac{q}{p}$ controls the shape, with $\frac{q}{p} > 1$ ensuring the s-shape of the cumulative diffusion curve. This implies that parameters *p* and *q* are negatively correlated. When eliciting data to inform the Bass model, correlated parameters will have an impact on the fit of the elicited data. The extent of correlation of the Bass model parameters and its implications for further analysis will be explored in Chapter 6.

The shape of diffusion when using the Bass model is illustrated in Figure 4.2 using parameter values for *p* and *q* (*p=0.0009* and *q=0.6995*) that were estimated for a biomedical device in the US (Gobok et al., 2009), and a hypothetical value for the market size *m*.

Figure 4.2 Illustration of the per period and cumulative diffusion curves estimated with the Bass model



The classic Bass model requires knowledge of uptake in the previous period to estimate uptake of any given period. The value of zero for uptake in the previous period results in the future period uptake to be zero, too. To estimate uptake in the first period after technology introduction, another estimation method is therefore required. The following approximation to the number of adoptions in the first period after launch has been used (van den Bulte and Stremersch, 2004, Mahajan et al., 1990):

$$N_1 = \frac{m(1 - e^{-(p+q)})}{(1 + \frac{q}{p} * e^{-(p+q)})} \tag{4.24}$$

where N_1 reflects the purchases in year 1.

Another known quantity that describes the Bass model and is useful for the elicitation exercise in Chapter 6 is the time period at which the inflection point of the Bass model s-shaped curve occurs (Mahajan et al., 1990, Sultan et al., 1990):

$$t' = \frac{1}{p+q} * \ln\left(\frac{q}{p}\right) \tag{4.25}$$

where t' is the point of inflection, or the time in number of years at which the number of adoptions starts to decline.

4.4.2 The Satoh approach to modelling diffusion

I highlighted in Chapter 3 that there are two potential problems with the use of the classic Bass model in an elicitation exercise to predict diffusion without any available data. The Bass model limitations were a) that parameters are correlated and this correlation is not captured when parameters are elicited univariately; b) the representation of discrete uptake data in a time-continuous fashion. These problems together may result in unrealistic diffusion curves. The alternative approach which was proposed by Satoh (2001) is a transformation of the Bass model that uses a discrete analogue of the Riccati equation proposed by Hirota (1979) to apply to the Bass model and make it discrete.

A Riccati equation is any quadratic first-order differential equation such as this one:

$$\frac{du}{dt} = a(t) + 2b(t)u + c(t)u^2 \tag{4.26}$$

where $a(t)$, $b(t)$ and $c(t)$ are given functions of *t*.

Hirota obtained a discrete analogue of the Riccati equation that has an exact solution for *u(t)* when *a*, *b* and *c* are constant over time (Satoh, 2001, Hirota, 1979):

$$\begin{aligned} \frac{u(t+\delta) - u(t-\delta)}{2\delta} \\ &= a + b\big(u(t+\delta) + u(t-\delta)\big) \\ &\quad + cu(t+\delta)u(t-\delta) \end{aligned} \tag{4.27}$$

with δ being the constant time-difference length.

Assuming *a*, *b* and *c* to be constant, Satoh (2001) set them equal to the following terms consisting of Bass model parameters:

$$a = mp \tag{4.28}$$

$$b = \frac{q-p}{2} \tag{4.29}$$

$$c = -\frac{q}{m} \tag{4.30}$$

Putting these back into the Hirota discrete analogue yields (Satoh, 2001):

$$\begin{aligned} \frac{N_{n+1}-N_{n-1}}{2\delta} = p\left(m - \frac{N_{n+1}-N_{n-1}}{2}\right) \\ + \frac{q}{m}\left(\frac{m}{2}(N_{n+1}+N_{n-1}) - N_{n+1}N_{n-1}\right) \end{aligned} \tag{4.31}$$

with $n = \frac{t}{\delta}$.

The solution to this is:

$$N_n = m\left(\frac{1-\left(\frac{1-\delta(q+p)}{1+\delta(q+p)}\right)^{\frac{n}{2}}}{1+\frac{q}{p}\left(\frac{1-\delta(q+p)}{1+\delta(q+p)}\right)^{\frac{n}{2}}}\right) \tag{4.32}$$

Reflecting that each time period is one year, the parameter δ can be set to $\delta = 1$ and Equation (4.32) simplifies to:

$$N_t = m\left(\frac{1-\left(\frac{1-(q+p)}{1+(q+p)}\right)^{\frac{t}{2}}}{1+\frac{q}{p}\left(\frac{1-(q+p)}{1+(q+p)}\right)^{\frac{t}{2}}}\right) \tag{4.33}$$

It is immediately obvious, that N_t now is independent of the number of adoptions in the previous period and therefore the Satoh approach enables estimation of the number of adoptions in the first period N_1 directly, without using an alternative estimation approach. The performance of the Satoh approach compared with the classic Bass model will be compared in Chapter 6.

4.4.3 Summary

The Bass model of diffusion and the adaptation by Satoh provide a structure for incorporating diffusion estimates in EVII and DCEA analysis.

4.5 Discussion

4.5.1 Findings

In this chapter, I proposed a modelling framework for the incorporation of diffusion estimates in health economic evaluation. These frameworks included the dynamic EVII model, the dynamic CEA model and the diffusion model.

4.5.2 Relevance

The dynamic EVII framework allows decision-makers to assess the opportunity loss to the health system that is caused by recommending a technology under uncertainty and imperfect implementation. As such, this framework can help direct investments to the most valuable activities, whether they be research studies or implementation measures. At the time this framework was developed it was the only one that allowed diffusion dynamics to be incorporated in these analyses. It remains the only one doing so using an established diffusion model with behavioural parameterisation. It also is the first to appropriately consider the relationship between implementation and information and its implications for decision-makers.

I have demonstrated within my framework that other authors' assumptions and interpretations used in and derived from value of implementation methods were incorrect. Since the time I developed this framework, one study appeared that claimed to consider the effect of implementation on the value of information (Andronis and Barton, 2016). This study had several limitations that I addressed with my framework. One of those limitations was that their proposed way of incorporating time-varying implementation estimates was simplistic, as they made assumptions of linearity. More importantly, the authors misunderstood the concept of the realisable EVPI developed by Fenwick et al. (2008) and

subsequently their interpretation of the implementation-adjusted EVSI was incorrect. I have provided proof that adjusting the EVSI calculation itself by implementation results in calculating the EVSIM of research. One reason that may have contributed to the misunderstanding is the term 'realisable EVPI' that characterised the maximum value of implementation that can be resolved through research. I therefore purposefully called this the Research EVPIM instead, to avoid this possible confusion.

The DCEA framework has the potential to help decision-makers assess cost-effectiveness of a technology that may experience price changes in the future. This can potentially prevent a situation in which a technology is rejected based on its current price, which would have become cost-effective in the future, in cases in which the price change is precipitated by the reimbursement decision.

4.5.3 Strengths and limitations

The strength of this modelling framework relates to the fact that it is based on diffusion theory and established diffusion models to incorporate the dynamics of implementation. Furthermore, the dynamic EVII framework establishes more clearly and hopefully more intuitively than it was done in the past what the true effect of information on implementation can be and how it can be calculated. The strength of the DCEA framework is that it formalises a phenomenon that exists (at least in some health technologies, namely medical devices) that has not been widely considered in the context of HTAs.

A limitation of this chapter is that the implementation modelling framework has not been applied in a case study. This is going to be done in Chapters 7 and 8. Another limitation is that only one diffusion model (and one adaptation of it) is considered in this framework. It was shown in Chapter 3 that other diffusion models are possible. However, the Bass model was chosen on the basis that it was the most widely used and established model, with predictive accuracy shown in many studies.

4.5.4 Conclusion

This chapter has provided a framework for incorporating diffusion estimates in health economic modelling for use in HTA decision-making. This framework can enable more informed decision-making and potentially direct investments to where the greatest value to the health system can be achieved. Further research is needed to apply this framework in a

case study with the objective to assess its feasibility and potential range of results and to explore any issues that may arise with its use.

CHAPTER 5. IDENTIFYING POTENTIALLY EFFECTIVE IMPLEMENTATION STRATEGIES FOR EIS IN PTB SCREENING

5.1 Introduction

One of the objectives in this thesis was to identify relevant implementation strategies for the case study technology Electrical Impedance Spectroscopy (EIS) for use in preterm birth (PTB) screening. The methods for estimating implementation have been reviewed in Chapter 3 and the model framework for the evaluation of implementation strategies has been developed in Chapter 4. To facilitate estimating the effect of an implementation strategy on the uptake of EIS in Chapter 6 and evaluate its cost-effectiveness in Chapter 7, relevant implementation strategies are identified in this chapter. The results will ultimately be used to calculate the expected value of implementation of the identified strategies. Figure 1.2 in Chapter 1 illustrates the flow of the thesis.

The rationale for the study presented in this chapter is that there is currently no knowledge of potentially relevant implementation strategies for EIS in PTB screening. As discussed in Chapter 1, it is not possible to use readily available strategies from other technologies because there is a wealth of factors influencing diffusion of health technologies rather than a well-defined set of main drivers. It is also plausible that different implementation strategies exhibit different levels of effectiveness at different stages of the product lifecycle.

The aim of this chapter is to identify relevant implementation strategies for the use of EIS for PTB screening in the England NHS at an early stage of technology life. Qualitative research methods that enable asking open-ended questions and provide the depth to explore why phenomena occur (Ritchie et al., 2003) were therefore deemed ideal for this study. This chapter is structured as follows: in Section 5.2, I describe the selection of the methods used, the sampling frame, interview design, data analysis and methods used to identify implementation strategies that are relevant at the current stage of the technology life cycle. I then present the results from the interviews in the Results Section 5.3, including the transferability of results and triangulation with the subsequent elicitation study and present the selected implementation strategies. I conclude with a discussion and conclusion in Section 5.4.

5.2 Methods

5.2.1 Method selection

In this section, I present how I identified potential implementation strategies in two stages. The first phase entailed an exploratory study to identify potential uptake determinants for the case study technology EIS. In a second step I chose implementation strategies relevant to EIS at its lifecycle stage.

5.2.1.1 Stage 1: Identifying potential uptake determinants for EIS using qualitative research

The research question in this part of the chapter was to identify potential factors influencing the uptake of EIS. It may not be possible to turn all uptake determinants into an implementation strategy. However, there were two reasons for keeping the research question this broad: first, the desire not to limit respondents unnecessarily in their responses. The more restrictive the design of the topic guide, the smaller the likelihood that respondents would think broadly and name relevant factors that do not come to mind easily. Second, so as not to miss potentially valuable information. While some named uptake determinants might not translate into implementation strategies, they might give a qualitative description of how best to design those strategies.

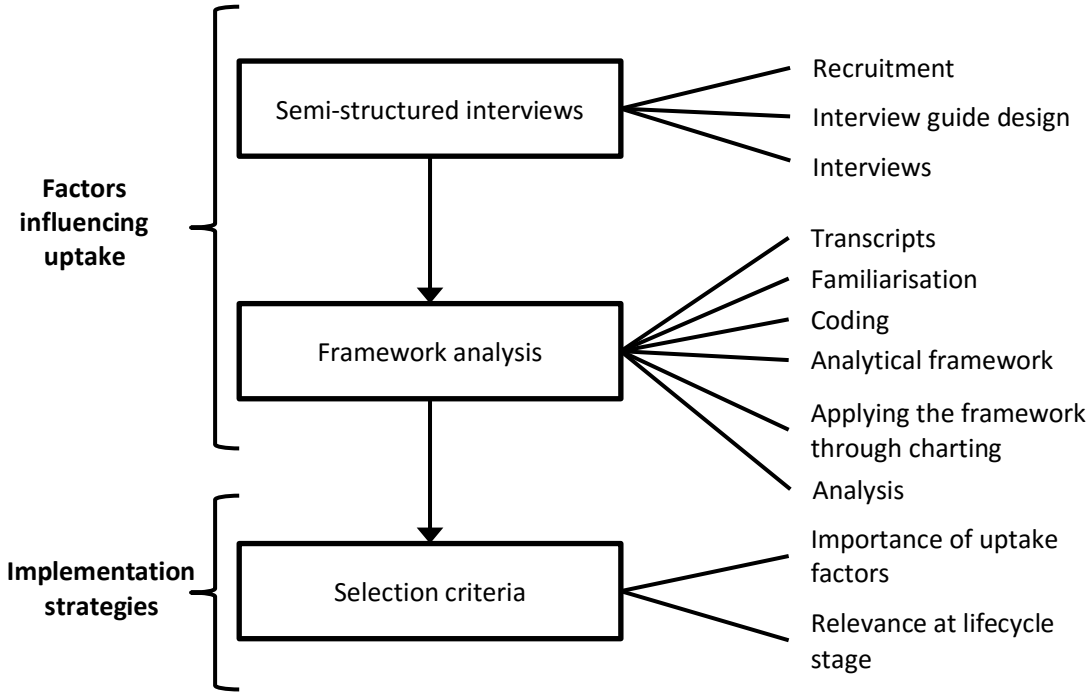
Semi-structured interviews with experts were deemed the most appropriate format to elicit information on determinants of uptake. Semi-structured interviews allow asking open-ended questions while providing a structure to guide the researcher and interviewee through the interview (Britten, 2006). I adopted a topic-centred approach, that is, there was an interview guide that covered certain topics to make sure these topics were discussed but it also allowed themes to emerge when they had not been considered previously (Britten, 2006). This method also provided the flexibility to probe about past points or more detail. Individual interviews with experts were chosen over focus groups to capture the full scope of individual experiences, including differences between different professional profiles and settings. The adoption decision to be investigated is likely to occur at the level of individuals, such as individual clinicians or business managers, and individual interviews are more likely to capture individual thoughts and nuances. Focus groups do not offer the same opportunity for eliciting individual thoughts (Ritchie et al., 2003) and I therefore considered them less relevant for this study. One advantage of focus groups over individual interviews would be the opportunity of reaching consensus among experts with regards to the most important

uptake factors. This advantage is, however, offset by the fact that individual additions of themes are less likely to be captured. In addition, focus groups are associated with more practical complexity because they require gathering experts together at the same time. Study participants may not be available at the same time, and travelling to a meeting location would require additional time from the experts; hence, individual interviews at the work place of the participant were considered the most appropriate method.

5.2.1.2 Stage 2: Selection of implementation strategies

The intention was to choose potentially effective implementation strategies that made sense, rather than investigating all possible implementation strategies. Ideally, one would want to elicit the effects of all possible implementation strategies and optimise the decision to achieve the optimum strategy in terms of effects on implementation and costs, but this was not feasible within the scope of this project, nor did it serve the main purpose of illustrating the use of the proposed extension to the expected value of implementation framework. I therefore do not claim comprehensiveness in the selected implementation strategies, nor do I claim that these are the best implementation strategies in terms of their impact or cost-effectiveness. I nevertheless made sure that relevant implementation strategies were chosen by basing the choice upon the findings from the qualitative study and then applying selection criteria to those findings. This process is shown in Figure 5.1 and described in more detail below.

Figure 5.1 Process of identifying implementation strategies for EIS



5.2.2 Sampling

The population of interest, or the sampling frame, was based upon the inclusion of those individuals that would be involved in the selection, purchase and use of new technologies in obstetrics at the Sheffield Teaching Hospitals NHS Foundation Trust, Barnsley Hospital NHS Foundation Trust or Rotherham NHS Foundation Trust. With those three trusts, both hospital types in England, that is, University Teaching Hopsitals (UTH) and District General Hospitals (DGH) were covered. The restriction to three trusts ensured feasibility and facilitated recruitment. The sampling approach adopted was purposive sampling to cover a relevant range of contexts or phenomena (Mason, 2002, Ritchie et al., 2003) that would enable reflection of different ranges of experience.

Professions and different types of hospitals were chosen as the categorising factors for this sampling frame because it was expected that different professions and professional settings might contribute different views on the adoption decision. The sampling categories included obstetricians and business managers from either UTHs or DGHs. Obstetricians would be using EIS and would be the most crucial in proposing new devices to purchase. Business managers on the other hand would provide further insight into financial determinants in the purchasing decision. It was anticipated that affiliations with the different hospital types might shape individuals' experiences and thoughts on adoption decisions. To

cover the different categories, an absolute minimum of six respondents was needed (Table 5.1).

Table 5.1 The sampling frame for the semi-structured interviews

Category	**Number of individuals**
Obstetricians (DGH)	At least 1
Obstetricians (UTH)	At least 4
Business manager (DGH or UTH)	At least 1
Total	At least 6

The required sample size for qualitative interviews is determined by the point at which theoretical saturation is achieved; that is, when no new themes or data emerge (Guest et al., 2006). For fairly homogeneous samples, six to eight interviews have been described as potentially sufficient (Guest et al., 2006, Kuzel, 1992). Consensus theory developed by Romney, Batchelder and Weller (1986) as cited in Guest et al. (2006) suggests that the degree of expertise among the study participants, which may be defined by certain qualifications or experience, can ensure homogeneity of responses. This also suggests that for expert interviews fewer respondents are needed than for interviews with novices. As Guest et al. (2006) argue, a homogeneous sample of experts, together with a relatively narrow objective, may cause saturation to be achieved sooner. In this case, the degree of expertise was regarded as high as the sampling frame was narrowly defined. The objective of identifying a shared perception of the most relevant uptake factors was seen as relatively narrow, which justifies the relatively small minimum sample of six participants. Despite these theoretical justifications, the degree of saturation was monitored after conducting a number of interviews by examining how many new themes still emerged.

The sample was generated through snowballing. Snowballing refers to a process of asking people who have been interviewed to identify other potential participants who match the selection criteria (Ritchie et al., 2003). Contacts with one person from each trust were established at the start of the process. Selection criteria were shared with the individuals that were willing to participate. These individuals then contacted potential participants and provided the details of those who might be interested in taking part.

Interviews with 10 experts involved in purchase decisions and use of obstetrics technologies were conducted between June and October 2013. These included obstetricians from both UTHs (7) and DGHs (2) and one business manager. All themes and sub-themes had emerged at the sixth interview, achieving theoretical saturation. Although there was a possibility that further themes would emerge with more interviews, the additional marginal benefit was deemed too little to continue the interviews as it was not expected that the overall results would change. I also considered the impact of a potential lack of saturation to be small due to the objective of this research to inform an elicitation exercise in contrast to developing a comprehensive understanding of uptake determinants.

5.2.3 Interview design

An interview guide was created to ensure that all interviews followed a similar format. The interview guide started out with broad questions and became more specific to ease the interviewee into the topic, as recommended by Legard et al. (2003). In the introductory part, background on the study was provided. An explanation of uptake of health technologies was provided. The information sheet was discussed (respondents had received it in advance of the interview) and any questions the participants had were addressed. The introductory part concluded with warming up questions about general practice.

This was followed by the main part in which the interviewees were asked about anything that might influence uptake within certain categories. To develop these categories I considered consulting existing catalogues of diffusion determinants by Greenhalgh et al. (2004), the DH report (Department of Health, 2011) and the qualitative study by Barnett et al. (2011), that were described in Chapter 1 and by thinking about what was going to be easiest for participants to think about, without prompting them. Most categories named in the Greenhalgh study were quite abstract and not intuitive enough to use as guiding questions in an interview schedule. The categories named by the DH report, on the other hand, were very concrete and were likely to prompt participants if included in the interview guide. These categories were also too narrowly defined, making assumptions on the effectiveness and cost-effectiveness of the technology in question. I therefore decided not to adopt any of these catalogues, but to use them as inspiration and design the interview guide based on categories that were intuitive enough to provide a guide for participants to direct their thought while attempting to cover a wide range of topics.

The final questions asked about factors influencing diffusion were related to: the technology; the users; the purchaser; the organisation; the external environment; and anything else that could increase uptake if low uptake was noted. I included the question on technology-specific factors influencing uptake because the Greenhalgh et al. (2004) study emphasised the role of the technological advantage in diffusion. Equally, the questions on the organisation and the external environment were included because of the importance placed on these in the Greenhalgh study, the qualitative study and the DH report. The other questions were designed to include all factors influencing different stakeholders in their decision-making (the user and purchaser). This design allowed for any thoughts to come up without prompting, with subsequently aiding the stream of thoughts by allowing the questions to be more specific and targeted. Finally, respondents were asked to name the two to three factors thought to be the most influential on uptake of EIS out of all those that they had mentioned. If required, all the factors they had mentioned before were repeated to them (this was made possible through taking field notes) but it turned out that most respondents had a very clear idea of what was most important in their opinion.

All interviews were conducted by me. This ensured familiarisation with the interviewees' different responses and picking up on nuances as well as learning about how best to approach different topics. Most interviews lasted 30-40 minutes, but the shortest was 28 minutes long and the longest one hour and six minutes. Interviews were audio-recorded and field notes of important points were taken. I transcribed all interviews and double-checked transcripts for accuracy.

A pilot was conducted with one respondent. This facilitated the review of the interview guide for its effectiveness in eliciting the information needed and served as practice in performing interviews as this was the first time I ever conducted semi-structured interviews. The interview guide was not changed posterior to the pilot interview and as a result of this I decided to include the findings from the pilot in the data analysis.

5.2.4 Ethical considerations

The study had obtained ethics approval by the University of Sheffield Research Ethics Committee (UREC), as well as NHS R&D governance by the Sheffield Teaching Hospitals NHS Foundation Trust, the Rotherham NHS Foundation Trust and Barnsley Hospital NHS Foundation Trust. Interviews were held face to face at the respondents' work place to ensure convenience and safety. Written informed consent was obtained from all study participants

after questions raised by respondents were addressed before the start of data collection. The information sheet, informed consent form and interview guide can be found in Appendix C.

5.2.5 Data analysis using framework

Data analysis was performed using 'Framework' in order to identify and categorise the factors influencing uptake of EIS that are hidden in the data. Framework is an analytical process with various structured phases (Ritchie and Spencer, 2002). It is a matrix-based analytic method and a way of using labels and categories to organise and analyse the data (Ritchie et al., 2003) whilst enabling the researcher to move between different levels of abstraction. This is done by using cross-sectional code and retrieve methods (Mason, 2002) by which a system of categories is first devised and then applied to the data. A study will thus have a thematic framework, with data organised into key themes and divided into sub-themes. Framework has been developed for use in policy research and since been used widely in health and policy research because of its pragmatic outlook that makes it particularly suitable to help answer policy questions (Pope et al., 2006, Gale et al., 2013). Framework was chosen over other types of thematic analysis because of its differentiating step of charting themes. These charts display themes in the shape of a matrix that allows analysis within cases and within themes. With the preservation of individual accounts, framework facilitates comparisons across individuals (Gale et al., 2013). This appeared to be particularly useful in visualising the data, and categorising it.

After interview data has been transcribed, framework comprises five main phases (Ritchie and Spencer, 2002) shown in Figure 5.1; the first phase is that of familiarisation. Having conducted and transcribed all interviews myself helped with the first steps of familiarisation; and listening to the tapes and reading through the transcripts twice aided the process. More familiarisation with the data occurred throughout all phases of data analysis while coding, sorting and charting the data. In phase two the thematic framework, consisting of recurring themes and ideas, was identified through examining the range of responses to questions. A numerical index was created at that stage, which referred to ideas or themes that were close to the original verbatim rather than imposing existing theoretical structure at this point (Ritchie et al., 2003). Themes were sorted and grouped under higher categories. Indexing, that is applying codes to the data, was conducted in phase two, by printing transcripts and marking phrases and sentences of verbatim with the respective index in the margins of the transcripts. After a first round of indexing, an initial framework was established (phase three) based on abstracting from verbatim to develop themes. The

framework was then revised based on a second round of indexing in which themes were added and others were grouped if there was some overlap. At this stage, 49 codes were created and applied to the data. The codes were aggregated into a number of four themes.

With the framework developed, phase four entailed charting data thematically. For each theme abstracted summaries of respondents' statements were chosen and presented in a chart with one row for each respondent and one column for each sub-theme and page references for each summary. I attempted neither to over-simplify nor to include too much data. Row heights and column widths were kept the same for each chart in order to facilitate easy comparison across respondents. The charts were reviewed and compared with the initial framework to check whether any data were missing. In phase five I returned to the aim of this study and described the range and nature of factors influencing uptake of EIS and identified patterns and connections. Associative analysis, that is finding links and connections between phenomena and explaining why those associations exist, was conducted as there was evidence of some themes appearing in clusters. For this, associations were explored as pointers to further analysis, as recommended by Ritchie et al. (2003).

It was tempting to include the recurrence or the frequency with which a factor was named as a measure of importance. As highlighted by Ritchie et al. (2003), frequency should not be presented as a primary finding as it does not have any statistical value. On the other hand, the authors also argue that frequency of occurrence should not be ignored but explanations be sought for this phenomenon. It was for this reason that counts of respondents mentioning a factor were presented at times and explanations sought when some importance was placed on those counts.

5.2.6 Validation of findings

The relevance of validity as a concept has been disputed in qualitative research (Lewis and Ritchie, 2003). It traditionally consists of internal and external validity. The former relates to whether the research explores what the researcher claims to be investigating and the latter to whether the findings are applicable to other parts of the population which makes it a part of generalisation (Lewis and Ritchie, 2003). The questions posed in qualitative research differ from those in quantitative research (Lewis and Ritchie, 2003) and could be represented better by exploring whether the respondents' perceptions were accurately reflected. Having adopted a pragmatist stance and using a mix of qualitative, elicitation and modelling techniques within this study, I considered this study to share features with mixed

methods studies. The question posed in a pragmatist mixed methods study is more that of 'transferability' than of generalisation or context (Morgan, 2007). Examining transferability refers to investigating which factors can make these findings applicable to other settings. One way of achieving that would be to compare the identified themes to existing research. In this case the themes were compared to the broader diffusion model components developed by Greenhalgh et al. (2004), the barriers to implementation identified by the Department of Health report (Department of Health, 2011) and diffusion determinants in the England NHS elicited in a qualitative study (Barnett et al., 2011). At the centre of this comparison were the following questions: Were named factors able to fall into the same model components? If not, are there any evident explanations for them lying outside the existing model? If they did not cover all of the model components, was there any evident explanation for that? Findings from this comparison were then used as pointers to the factors that could influence the transferability of results to other settings.

Multiple assessment, that is, assessment by another researcher to ensure that themes were identified correctly, was considered but not used. There were a few reasons for not using it: there are limitations with this because no individual interview will match the abstract framework fully (Mays and Pope, 2006). As such, multiple assessment is not advised as a check on validity but rather as a type of error reduction in each interview (Mays and Pope, 2006). Another reason was that a degree of subjectivity is accepted in mixed methods research (Morgan, 2007) and that researchers outside this project would not know the data and meaning as well as me. Instead, transparency in analysing the data can help the reader in judging the way that conclusions were drawn (Mays and Pope, 2006). Framework helps in transparency by charting data from the interviews within the identified themes.

Further validation was performed through triangulation. Triangulation refers to comparing results from different methods of data collection and enables testing the comprehensiveness of findings (Mays and Pope, 2006). Hence, the qualitative explanations obtained in the elicitation study were compared with the findings of the present qualitative study to check for any deviations in what respondents regarded as important influencing factors to uptake. For this, all qualitative observations from the elicitation study were compared with all themes from the qualitative study described in this chapter.

5.2.7 Identifying potential implementation strategies

With factors influencing the uptake of EIS identified through semi-structured interviews and framework analysis, I turned to the second objective of this chapter: identifying potentially effective implementation strategies that could be assessed for their effectiveness in the subsequent elicitation study. This part of this chapter does not claim to identify the most effective or cost-effective implementation strategies – this could only be achieved through quantitative analysis of all of them. However, I attempted to identify implementation strategies relevant for EIS at the current stage of product lifecycle; that is, before its introduction.

I first reviewed answers to the final question asked in the interviews. In that question, participants were encouraged to name three factors they thought to have the greatest influence on uptake. It was assumed that ease of recall in this particular question truly reflected the importance of the factor. Second, I re-examined the language used by respondents and printed in the framework charts to identify what was described as vital or most important by respondents to check for any divergence from the respondents' identified uptake factors. Third, I placed greater emphasis on those uptake factors likely to be relevant at an earlier stage of product development. A chart of factors influencing the uptake of EIS and their relevance in the technology life cycle was therefore developed based on the different stages of product lifecycle (Kotler and Keller, 2012) and the respondents' statements.

5.3 Results

In Sections 5.3.1 to 5.3.4, themes are described in detail including their sub-themes and charts of representative summaries of respondents' comments belonging to each included theme and sub-theme are shown in Appendix C. The transferability of these findings is discussed in Section 5.3.5 and findings are triangulated with those of the elicitation study in Section 5.3.6. In Section 5.3.7, potentially effective implementation strategies are presented.

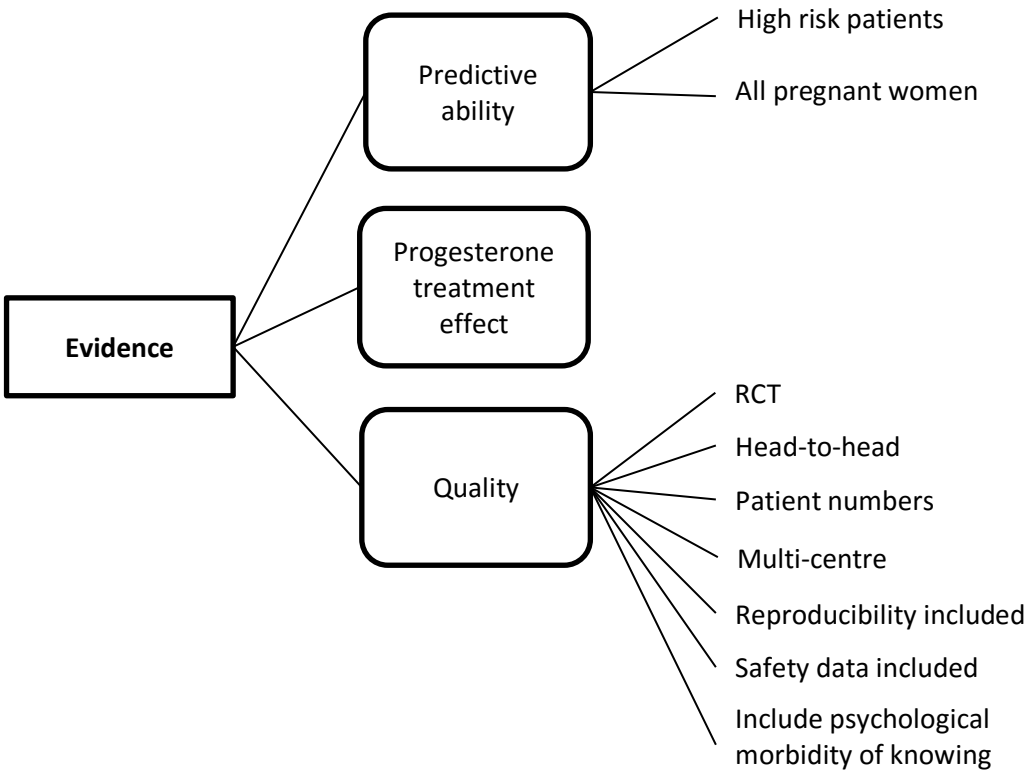
5.3.1 Research evidence

Research evidence was identified as a key theme with three sub-themes (Figure 5.2) and representative summaries of respondents' comments relating to its three sub-themes of

evidence for 'predictive ability of EIS', evidence for the 'progesterone treatment effect' relative to no treatment after screening with EIS and the 'quality' of the evidence are shown in Appendix C.4.

Evidence for the predictive ability of EIS was mentioned as a crucial factor by all respondents and evidence for the progesterone treatment effect was mentioned separately by all but one respondents. All but two respondents also mentioned quality of the research in some shape or form as a potential determinant to uptake. These three sub-themes included statements on its different aspects which are depicted branching out from the sub-themes in Figure 5.2. Each of the sub-themes with their different aspects are now described in further detail.

Figure 5.2 Theme 1: Research evidence



The predictive ability of EIS was also named as evidence *'that it works'*. Most interviewees thought that the predictive ability of EIS would be demonstrated in a study among high risk women but some others emphasised the desirability of showing predictive ability in the wider population of all pregnant women. This would enable a change in the

general perception that a predisposition to giving birth preterm can only be discovered in women with risk factors. Those risk factors entail a history of previous preterm births or miscarriages. It would obviously be desirable to prevent premature births also in women who have not experienced losses or premature deliveries of babies before.

> *'If it works we could screen all pregnant women at their 20 week scan.'* (Participant 1)

> *'There has to be overwhelming evidence, e.g. likelihood ratios in the 10s and 20s and good reproducibility.'* (Participant 2)

> *'...without trial evidence noone is going to introduce anything.'* (Participant 3)

> *'Research has to show sensitivity and specificity to make sure that clinicians will use it correctly.'* (Participant 8)

> *'The biggest thing would be the evidence, that it works, and if there was clear evidence on what the true positives, negatives and false positives, negatives were, that could drive uptake.'* (Participant 10)

If EIS could be shown to reliably predict PTB also in women without a history, then this would increase its utility to clinicians since such a tool is not available at present. Whether EIS would be used in a high-risk population only or in the wider population of all pregnant women would of course have huge implications for its absolute uptake. On the other hand, screening every pregnant woman, including women who would otherwise remain in midwifery-led care, would have clinical pathway and financial implications.

> *'The ability for it to be used by all pregnant women would be crucial for uptake.'* (Participant 1)

> *'If the evidence came out that you would be screening every pregnant woman, that would have quite big training and logistical issues.'* (Participant 4)

> *'If you start using it in high risk groups, will it translate to low risk groups, will it get false positives, so I think you need a lot more studies.'* (Participant 5)

The sub-theme of evidence on the progesterone treatment effect was also expressed as the desirable outcome of delivering *'happy healthy babies'*. The rationale behind this was that participants questioned the effectiveness of progesterone treatment, especially in combination with EIS screening. The evidence base for treatments in general is relatively immature, with progesterone therapy deemed to be the most promising treatment available at present (Honest et al., 2009). However, existing evidence only shows progesterone effectiveness in preventing PTB after cervical length scans. More studies to show the effectiveness of progesterone after EIS screening were desirable from the point of view of some respondents and some stated that trials on the prevention of PTB with progesterone treatment after testing with EIS were needed.

> *'If you see that it saves lives, and you get results in clinical practice.'* (Participant 3)
>
> *'Happy healthy babies is a valid endpoint.'* (Participant 4)
>
> *'If it gives us better health outcomes, it will be taken up quickly.'* (Participant 6)
>
> *'Other people are not convinced that the existing treatment works and may need more evidence.'* (Participant 4)
>
> '*If you could prove that there was an intervention to improve the overall outcome, then that would send up the uptake.'* (Participant 10)

The summaries in Appendix C.4 highlight that respondents placed great importance on both types of evidence, using key phrases such as *'the main thing…', 'the biggest thing…'* and making judgement statements about uptake such as *'… it may not be adopted'* and *'…that would send up the uptake'*.

The third sub-theme was quality of the evidence. Any research study should ideally fulfil different dimensions of quality. The robustness of the study design was mentioned as a key factor, with a randomised controlled trial considered the gold standard and sufficient patient numbers to power the study emphasised by a few respondents. A double-blind study was mentioned as desirable, albeit difficult to implement with EIS. Respondents wanted initial research to be validated in multi-centre studies across different settings and some mentioned that evidence on the performance of EIS compared with a comparator would be desirable. A few emphasised that they would want the reproducibility of results to be shown by the research. By reproducibility of results respondents meant a demonstration of achieving the same results if the same person was screened twice on the same day.

'It'll need validation in multi-centre studies across different settings.' (Participant 1)

'Evidence in the shape of randomised-controlled trials, double-blind in fact, is essential.' (Participant 5)

'Good research could convince me of whether this could make a difference. It has to be a good-sized study with adequate power. And it needs to be randomised. I'm not bothered about lots of trials as long as it's good quality and it's big enough and not biased in any way by the people who are trying to make it.' (Participant 9)

Respondents would also like to see whether screening with EIS was safe and well-tolerated. One respondent mentioned that it would be important to consider the psychological morbidity that may be associated with screening of pregnant of women.

'Actually, there is a morbidity associated with this, putting patients under some degree of stress and telling some of them that they may have a PTB, a psychological morbidity that it could ruin your pregnancy.' (Participant 8)

5.3.2 Product and service characteristics

The second key theme identified was 'Product and service characteristics', which included the three sub-themes of 'easy to use design', 'costs' and 'marketing' (Figure 5.3). The ease of use of the device was important to the majority of the respondents and was mentioned as one of the major factors to uptake. A design that could facilitate ease of use comprises many different aspects. Disposable probes were one such aspect, as it was not deemed feasible to send the device for sterilisation after every use.

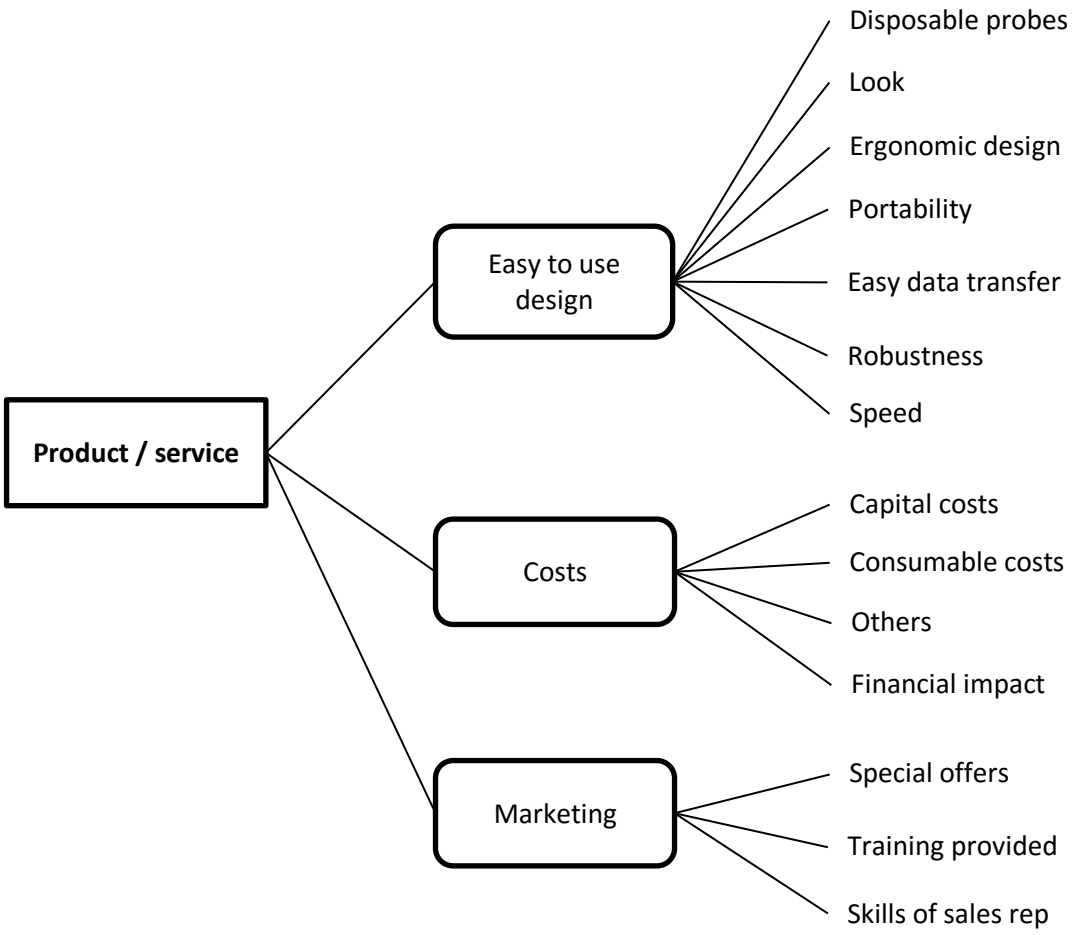
'It needs to have a disposable cover.' (Participant 4)

The look of the device was important to the respondents, who did not want it to look too '*menacing'* and thought that a sleek design would encourage usage and therefore uptake.

'It must not look menacing or too big, if it looks sleek and nice it is likely that people would take it up.' (Participant 1)

'The design is probably quite important, it should look sleek and easy to use and not big and threatening.' (Participant 5)

Figure 5.3 Theme 2: Product / service characteristics



Equally, it was important to respondents that the device be ergonomic in the hand of its user as an *'ergonomic design is a big factor to uptake'*. An ergonomic design was seen to entail a small size which would make it easier to handle.

'I've had issues in the past with devices that were too big and did not fit my hand.' (Participant 4)

A smaller size would also make it more portable. Portability was mentioned as important because that would enable its operation in different rooms without having to purchase a large number of devices. Another design feature that would make it more portable was the

preference for a battery-driven device rather than one that has to be connected to a charger overnight.

> *'Portability would influence the decision.'* (Participant 7)

Four respondents also made statements regarding 'easy data transfer'. The desirability of wireless data transfer to connect with any computer was mentioned along with the possibility of having a docking station for the device from which data could be transferred.

> *'It would be great if it had a docking station and you put it in and it downloads data to the computer.'* (Participant 4)

> *'If you've got a probe that will link to a computer by bluetooth or wifi, people think that's wonderful.'* (Participant 6)

The robustness of the device was highlighted by one respondent who thought that purchasers may be put off by a device that breaks easily.

> *'I'd want to know how robust is the piece of kit.'* (Participant 6)

Finally, respondents saw an advantage in EIS in that it could provide quick measurements and save time over other types of screening such as ultrasound scans. It was thought that speed of measurement would help bring it *'into clinical practice'*.

> *'If it's easy and quick to use it would influence uptake.'* (Participant 5)

The sub-theme of easy to use design was mentioned together with evidence by some respondents as in below example statements. This association of evidence and ease of use seems to highlight the relevance of those particular factors to clinicians.

> *'As long as the evidence said it was good and it was good in everybody's hands because it was easy to use, then I can see the potential.'* (Participant 9)

'An ergonomic and easy to use look is the greatest factor [to uptake], along with the evidence for its use.' (Participant 10)

The other sub-theme was the 'costs' associated with the purchase of EIS. Different types of costs were thought to have an effect on uptake. Based on the participants' statements it appears that the most important types of costs were the capital costs, that is the cost of purchasing the device, and the consumable costs, that is the cost of the disposable probes.

'When things are expensive they are not taken up…' (Participant 6)

'How usable it was would depend on the initial outlay…' (Participant 9)

'It would make a big difference if you had to think very carefully about consumable cost.' (Participant 5)

One respondent mentioned that the purchasing price mattered because with a price higher than £5,000 the decision would be made by the capital investment team, rather than the local directorate. It was also thought that if the cost of the disposable probes was too high, that is *'10s of pounds, [for example] £30 every time'*, then clinicians would think carefully about the use of EIS in clinical practice. One respondent mentioned *'that some companies will work with you to achieve that [low consumable costs]'*.

Other costs that were mentioned included training costs, replacement, maintenance, manpower and implementation costs. All respondents mentioned some type of costs and they placed some importance on them. Some stated that, in the current economic climate and given the NHS' reluctance to change, costs presented a very important factor.

'Cost is the biggest thing at the minute because all trusts are under pressure.' (Participant 9)

Because costs were placed in context with the effectiveness of the device or the wider financial impact the device could have, the financial impact was another aspect to the 'costs' sub-theme. Many participants mentioned that the cost of EIS would have to be outweighed by the savings or the benefit EIS could contribute in clinical practice and in terms of health outcomes.

'Why should we spend X amount of money, if we are not going to improve efficiency?' (Participant 2)

'If something can save money as well, the business case is a lot easier.' (Participant 3)

'The biggest thing would be evidence and working out the cost benefit, if you could save 3-5 PTBs per year, the impact that would have on drugs, special care baby units and long-term effects.' (Participant 10)

One respondent associated the cost-benefit of EIS with the use of EIS on all pregnant women:

'... if we had a simple device that costs just £5,000 that we could use on everybody and predict the ones that deliver before 28 weeks, then we could have some impact.' (Participant 4)

Several other aspects of the product and service characteristics were mentioned that were grouped under the sub-theme of 'marketing'. Within this sub-theme, it was mentioned that reduction of costs through training or maintenance packages offered for free by the manufacturer could have a positive effect on uptake. Trial periods in which the device is provided for free were seen as an initiative likely to increase uptake.

'It would have to be minimal outlay or it would be for the company to say we will give you this on trial, completely free, with X many refills for 6 months...' (Participant 2)

'It's how good care the manufacturer gives, that is also things like training packages.' (Participant 9)

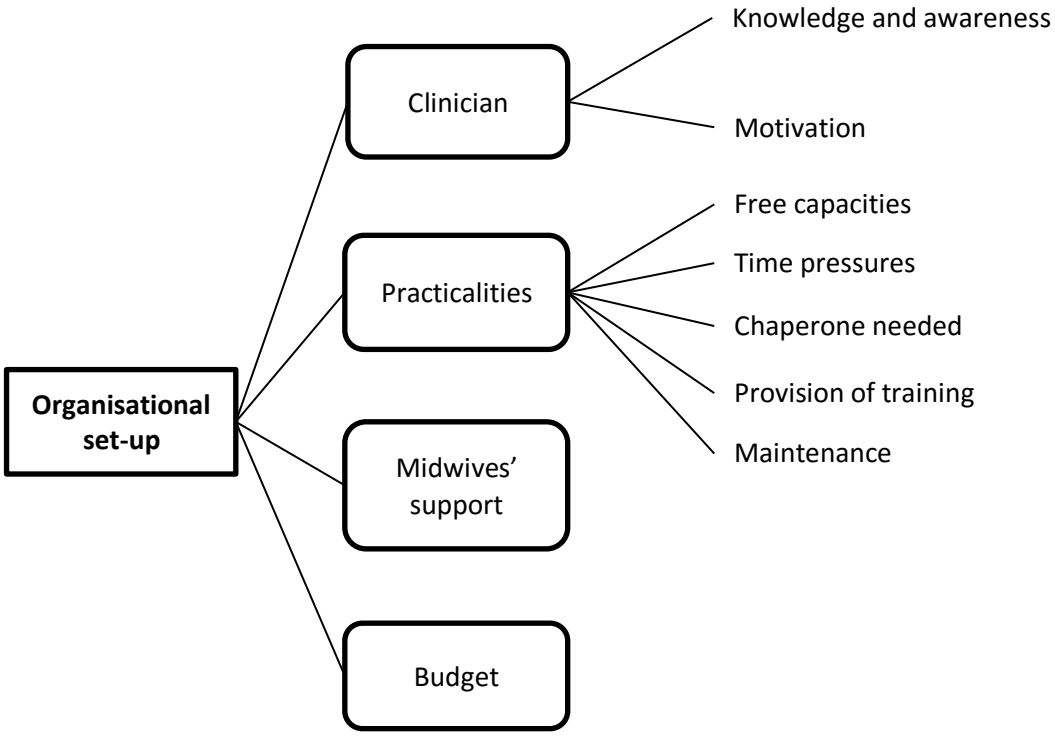
Furthermore, the skills of the sales representative, their enthusiasm and persuasiveness were seen to affect the uptake decision by individual clinicians.

'I think it's how enthusiastic the sales person is, if they believe in what they do and they can put that over to you...' (Participant 6)

5.3.3 Organisational set-up

The theme of the 'organisational set-up' contained the sub-themes of 'clinician', 'practicalities', 'midwives' support' and 'budget' (Figure 5.4).

Figure 5.4 Theme 3: Organisational set-up



There were two aspects to the 'clinician' theme: their knowledge and awareness and their motivation. Clinicians' potential ignorance of new evidence was mentioned as a potential barrier to adoption. It was also mentioned that some gynaecologists working in obstetrics may be less aware of recent developments in obstetrics and that junior staff on rotation may not be aware of procedures.

'It's got mostly to do with awareness of how good the technology is.' (Participant 9)

The motivation aspect emerged in eight interviews. It was never mentioned to be a very important factor but interviewees did seem to think that it could affect uptake. Respondents described enthusiasm for innovations or the lack of it as well as the effect that the career stage or age could have on the clinicians' attitude towards new technologies. The burden of

increasing workload through adopting new products or processes was also mentioned as a deterrent factor to uptake.

> *'When you have come to a certain level of training or age then it is very difficult to go for something new, it's treated with suspicion.'* (Participant 2)

> *'When new research recommends not to do something anymore, everyone immediately stops doing it, but when you're supposed to do something in addition to what you're already doing, uptake is a bit smaller.'* (Participant 4)

Among the sub-theme of 'practicalities', the ability of EIS to free capacities was seen as a major facilitator to uptake as five respondents emphasised the existing pressure on ultrasound scanning and that anything to relieve that would be welcome. They also thought that EIS may be more efficient than ultrasound or offer some cost-savings compared to ultrasound.

> *'Anything that will reduce the pressure on ultrasound will be welcome.'* (Participant 3)

> *'If you do it instead of [ultrasound] scans, then this is much quicker and there's efficiencies to be had.'* (Participant 4)

> *'Ultrasound seems to be such a scarce resource that any alternative would be welcome.'* (Participant 7)

Several other aspects were aggregated under the 'practicalities' theme: the need for training certain personnel, having a chaperone when performing the exam, and the time pressure caused by the additional work that would need fitting in with consultants' schedules.

> *'You can't do anything now, without someone being on a training course.'* (Participant 2)

> *'You may need someone to chaperone you and that can be an issue in an antenatal clinic.'* (Participant 5)

> *'If I need to do it to every patient that I see in antenatal clinic, that is going to add 15 minutes and reduce my clinic by a third.'* (Participant 6)

Midwives' support was mentioned and that getting them on board could have an effect on uptake.

> *'The midwives are often the ones who see it, they'll suggest why don't you do this test, so I think it's important to engage them.'* (Participant 5)

The last sub-theme within the organisational set-up was the organisation's budget that may prevent EIS from being purchased.

> *'Budget is an issue, for example trusts or health funders might decide to buy 4 rather than 8 devices.'* (Participant 1)

It was furthermore mentioned that there are different budgets for different purposes and that there is a certain flexibility as to what budget might be used:

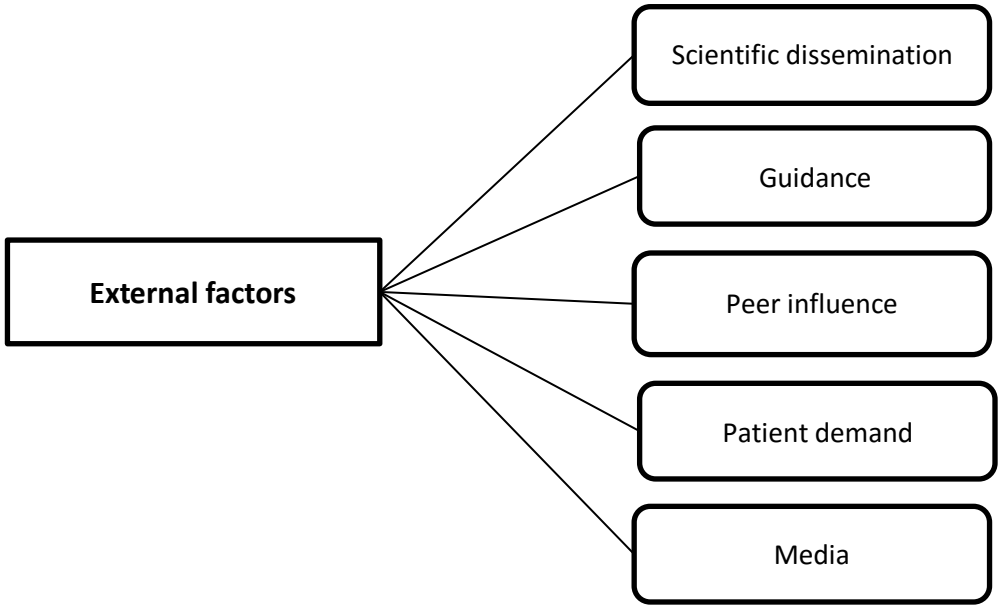
> *'It is about whether it is capital or revenue budget. We could be totally broke and load the costs at the front end because it's not our directorate's budget, but it's still the trust's budget.'* (Participant 7)

One respondent mentioned the fact that tariffs often only poorly reflect the actual costs in a directorate and where those costs occur (participant 7). The existence of perverse incentives was noted, where a less invasive lower cost procedure leads to a lower tariff despite it being difficult to reduce the actual costs associated with this procedure.

5.3.4 External factors

The theme of 'external factors' comprised five sub-themes: 'scientific dissemination', 'guidance', 'peer influence', 'patient demand' and 'media' (Figure 5.5).

Figure 5.5 Theme 4: External factors



Within the sub-theme of 'scientific dissemination', respondents placed importance on publicising research in different channels, e.g. in scientific meetings.

> *'Research needs to be publicised and the more it is, the more likely people are to think that this might be helpful.'* (Participant 4)

The sub-theme 'guidance' was mentioned by all respondents. They referred to different types of guidance such as those issued by local networks, the Royal College of Obstetricians and Gynaecologists, the national screening programme and NICE. NICE guidelines were the ones mentioned most often and a few times with a special emphasis on the ability of NICE guidance to make a significant difference to uptake, for example:

> *'The essential factor for the diffusion of a device is the reception that the NICE gives to it.'* (Participant 3)

Respondents mentioned that it was difficult not to adopt something once there were guidelines of any type and justification would have to be provided.

> *'If the Royal College or NICE said everyone must have this then it would be difficult to say No.'* (Participant 4)

The importance of ‘peer influence’ and actions of other hospitals was also emphasized:

> *‘Bad experience elsewhere can bring uptake down.’* (Participant 6)

> *‘The greatest rise in uptake will be from people who just follow the trend.’* (Participant 10)

Patient demand was also seen to be vital for uptake. This could go in both directions, with patients not accepting EIS, thus hindering uptake or patients demanding screening for PTB, thus encouraging the uptake of EIS.

> *‘If patients don't like it, it won't get taken up.’* (Participant 6)

> *‘A lot of bits of equipment are bought because patients in support groups talk about it, ask for it and want to go somewhere where it's available.’* (Participant 5)

The last theme was ‘media’ which, according to a few respondents, could have a significant influence through driving patient demand but could also highlight any potential problems associated with the use of a technology.

> *‘User feedback can be a big influence and it can come from women's magazines, soft press stuff and patients. It's getting it out there.’* (Participant 6)

> *‘The media can have a strange effect, in front of a paper would do wonders but it also picks up problems.’* (Participant 10)

It should be noted that sub-themes belonging to the ‘external factors’ theme were often mentioned together. This follows from two issues: firstly, that there was a question specifically asking for external influencing factors to uptake and secondly, that only very few respondents mentioned any theme belonging to that category before that question came up (it was one of the last questions). While this suggests that respondents were prompted, I believe that there is validity to these responses nevertheless. This is because respondents did seem to realise that these factors had slipped their minds before, and they often highlighted

them as being among the most important drivers to uptake. This shows the complexity of the topic and the difficulty respondents had in calling to mind all the relevant factors.

5.3.5 Transferability of results

To put results into the context of other research and obtain pointers as to whether the present study findings are transferable, the study findings were compared with studies of Greenhalgh et al. (2004), the NHS barriers to implementation described in the report by the Department of Health (2011) and findings from a qualitative study (Barnett et al., 2011). Greenhalgh et al. (2004) describe nine major model components with many sub-categories that are the result of a synthesis of empirical and theoretical findings. As such, their model is much broader and it is not surprising that all the included themes from this study fit into the diffusion model presented by Greenhalgh et al. (2004). Apart from being more limited in numbers, this study's themes differed from themes in Greenhalgh et al. (2004) in the way they were categorised. This is largely because a much more context-specific research question was used in this study. A comparison is provided in the following.

Sub-themes of the themes 'evidence' and 'product & service characteristics' can be associated with the Greenhalgh model component of 'Innovation', more specifically the sub-components of 'relative advantage' and 'complexity' (Greenhalgh et al., 2004). Sub-themes related to the 'organisational set-up' would mainly fall into Greenhalgh's 'implementation and routinisation' component, such as 'organisational structure' and 'intra-organisational communication' although they could also be ascribed to other areas such as 'system readiness for innovation' (Greenhalgh et al., 2004). My themes of 'external factors' would fall into their 'outer context' (Greenhalgh et al., 2004). Many other of the Greenhalgh model components were not covered by the findings of this study. The following reasons for why not more of the Greenhalgh model components were covered came to mind. First, Greenhalgh et al. (2004) themselves assert that their model is not a prescriptive formula but rather a memory aide for considering the different aspects of diffusion. Second, every technology has a different set of features and will be employed in different contexts, thus making some of the Greenhalgh model components redundant for individual technologies.

The present study findings can also fit with the diffusion facilitators identified by the Department of Health (2011). Those consisted of top-down, horizontal and bottom-up pressures and included guidance, skills development, incentives, peer influence, marketing, competition and patient demand, to name a few that predominantly match my identified

themes. The fact that sub-themes falling in the areas of 'evidence' and 'product & service characteristics' were not mentioned in the DH report, can be explained by its focus on existing innovations that have been shown effective but exhibit low implementation. As such, the DH report focused on a later stage of the product lifecycle than this study.

The qualitative study on diffusion determinants in the NHS (Barnett et al., 2011), in contrast, featured 'evidence' as a main theme. The study by Barnett et al. (2011) also prominently featured 'contextual factors' that could be intra- or interorganisational, similar to the themes 'organisational factors' and 'external factors' identified in this study. The Barnett theme 'people-based resources' was alluded to in my sub-themes 'midwives' support' and the 'clinician', although Barnett et al. (2011) mentioned a wider range of different people, including people in the local community and the senior management. These slight differences are owed to differing contexts of these studies. Lastly, Barnett et al.'s theme of 'the role of partnerships' was not extensively covered in this study. The sub-theme 'peer pressure' in this study touches upon this but other than that, partnerships were not viewed as vital by the participants in this study. It is difficult to assess whether this disagreement is owed to the context, the participants or the methods used in this study.

As was expected, study findings overlapped to a certain extent but not fully, due to different study contexts and potentially the methods of eliciting the relevant factors and the choice of participants. Despite this, I believe that study results captured well the potential barriers and facilitators to uptake of this particular technology and purpose and at this point in time. There are three arguments to back this up. First, the sampling frame which included interviewees from different hospital types and from different professions has been vital to covering different experiences. Second, the semi-structured topic guide was flexible enough to allow themes to emerge but pointed respondents in directions of where to search for possible factors when they did not come to mind easily. Third, framework analysis then ensured that all the themes were identified and categorised in a transparent manner.

However, there are three restrictions to the transferability of results:

1. Some of the uptake factors are certainly transferable to other technologies, but with the caveat that a technology-specific study such as this one cannot be a reliable source for informing diffusion determinants in other technologies. For example, it seems intuitive that there has to be strong evidence for any type of health technology, while other themes such as the easy to use design may be less applicable to other technologies. Also, the strength of a qualitative analysis such as this one is the detail it provides. While strength of evidence will be a factor for most health technologies, this study has provided a detailed description of

what evidence should look like; including potential endpoints and other features of study design that may not be transferable to other settings.

2. The results may not be fully applicable for the same technology with a different purpose. A similar technology to this one already exists for cervical cancer screening. Again, the study results presented here may partly be transferable to that setting but it seems safe to assume that other factors would play a role as well.

3. This study captures a snap-shot of the relevant uptake factors for EIS. Performing the same interviews at a later stage of product life, for example when EIS has already been available for a few years and the evidence base is greater, may bring up different factors. This is for three reasons: a) the setting in which the technology is employed may genuinely change, that is there may be new evidence, competitors, tariffs, budgets or guidance, amongst others. b) Some of the themes might not come to the minds of respondents anymore if they think that they are fulfilled; for example, they may deem the evidence base sufficient. c) The product itself may have changed and may, for example, have become more effective, cheaper or easier to use. Hence, even though the interviews covered all the factors that may be relevant at any point in time, they were based on current knowledge.

5.3.6 Triangulation

Triangulation with the results from the elicitation study revealed that the qualitative study had uncovered most of the relevant factors to uptake. Sub-themes belonging to all four themes were mentioned by experts in the elicitation study, with 12 out of 15 sub-themes mentioned. Unsurprisingly, the most important uptake factors identified from the interviews came up in the elicitation but there were also two new aspects that had not been mentioned before. One of these aspects was that respondents felt that the process of making a business case takes a long time which would delay the uptake of EIS, albeit not change its absolute size. This came up in the elicitation study because there was a specific question about timings of uptake which was not the case in the interviews. The other aspect was the size of the hospital or of the obstetrics unit that may have something to do with whether EIS would be adopted or not. This may mainly be related to the budget of such smaller units as well as the facilitation of the service (for instance, an arrangement could be made that high risk women could be treated in larger units only). At present, the way the service would be facilitated is unknown, but the finding is still relevant when estimating the total number of units that may adopt EIS.

5.3.7 Generation of evidence identified as effective implementation strategy

Examining answers to the interview question about those factors that would have the greatest impact on uptake and the language that respondents used throughout the interviews, narrowed the pool of most effective uptake determinants down to ten factors, described in the following. All interviewees stated that generation of further evidence was one of the factors with the greatest impact. While some interviewees viewed evidence on the predictive ability of EIS as most important, others emphasised the relative treatment effect of progesterone compared with no treatment. The use of language confirmed the perceived importance of strong research evidence to diffusion throughout the interviews: evidence on the predictive ability of EIS and on the progesterone treatment effect were both mentioned by most respondents in phrases that contained words such as *'crucial', 'the main thing'* and *'the biggest thing'*.

Other factors mentioned in response to this question mainly fell into the 'costs' and the 'easy to use design' sub-themes, with five respondents each mentioning these. 'Financial impact' and 'peer influence' were mentioned by four and two respondents, respectively, and 'clinicians' and 'patient acceptability' and the 'applicability to the wider population of pregnant women' were each mentioned once. Some of these were also associated with the use of superlatives. Especially participant 10 emphasised the importance of several different factors by using phrases such as *'the greatest factor'* referring to an *'easy to use design'*; *'the biggest thing'* referring to 'evidence and the cost benefit' and *'the greatest rise in uptake'* being caused by 'peer influence'. Unsurprisingly, participant 10 mentioned evidence, cost-benefit and ease of use as the most important uptake factors.

Factors that were not mentioned in response to the final question but came up as very important throughout the interviews were the 'clinician's awareness', that could *'influence its use in practice more than anything else'* (Participant 5) and 'guidance', for instance by NICE, that was described as the *'essential factor for the diffusion of a device'* (Participant 3).

These ten factors fell into the seven sub-themes of 'easy to use design', 'evidence on predictive ability', 'evidence on progesterone treatment effect', 'costs', 'guidance', 'practicalities' and 'clinician'. As I could not assess the value of all of these through elicitation due to time constraints, I had to make a final choice on the implementation strategies to evaluate in Chapter 6. For this, I based the decision on the relevance that these factors had at an early stage of product life. I developed a diagram that shows at which points of the product lifecycle, as presented by Kotler and Keller (2012), the potential implementation strategies would have the greatest effect (Figure 5.6). For this, I placed the

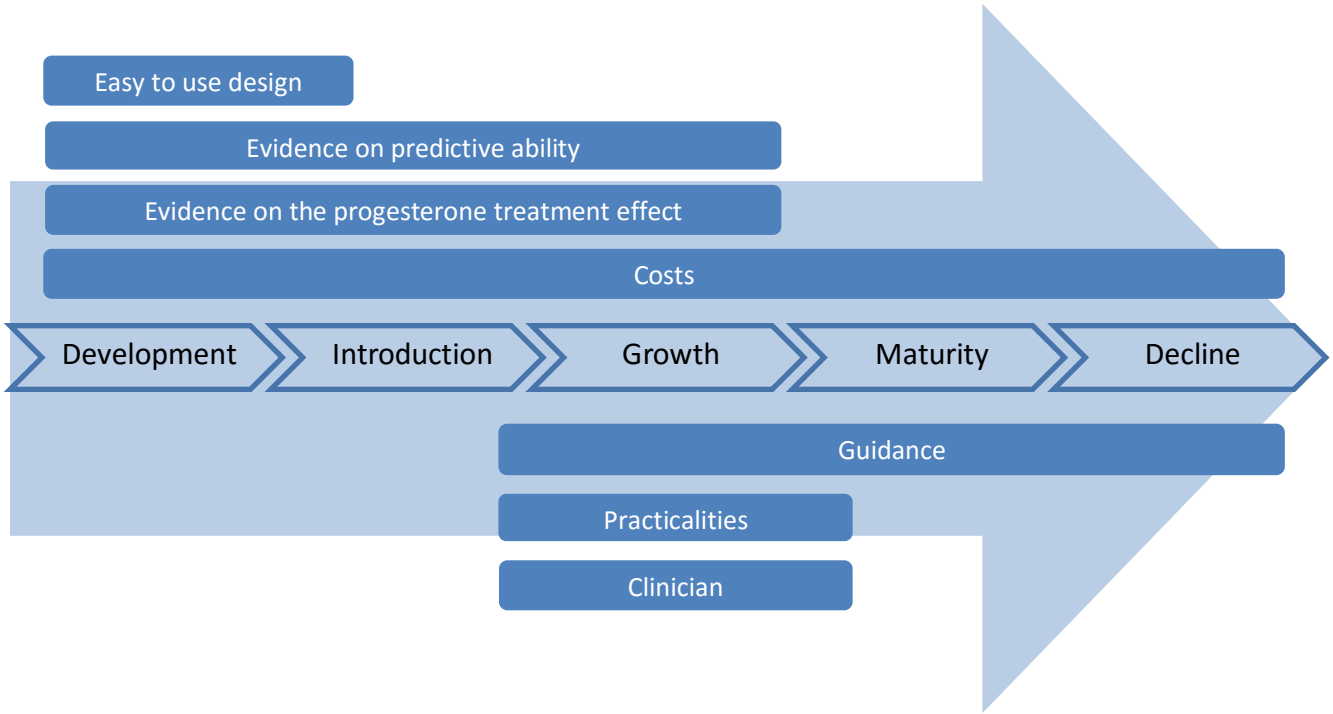
sub-themes identified above along the lifecycle arrow where I thought them to be the most effective. The limitation of this diagram is that the lifecycle stages are arbitrary in their shape and their duration, a common criticism of product lifecycle diagrams (Kotler and Keller, 2012). It follows that the placement of the uptake factors along the lifecycle stages is also arbitrary. However, the diagram illustrates what uptake factors may be targeted first in order to increase implementation of EIS throughout its lifecycle.

According to the interviewees, evidence on the predictive ability of EIS and the progesterone treatment effect would have to be right first. Unless the evidence was right and of high quality, other factors such as guidance or the ease of use would not significantly influence uptake and I therefore thought that getting the evidence right has to be made a priority in the early stages of life; that is, pre-introduction up to shortly after introduction. Of course evidence that emerges later on will still have an impact on uptake. At the same time as developing the evidence, the technology developer can adapt the design of EIS to make it more user-friendly and thus influence uptake. Again, the technology can be developed further at any stage of its life, but any actions would likely be more effective if they were taken before technology introduction. Costs can also influence uptake throughout the product lifecycle. Guidance and activities to address practical issues within the organisation will be developed most likely after technology introduction in the growth phase of EIS, as they are contingent on evidence being available and the product development to be more or less concluded. The clinician's awareness and motivation would also most likely be addressed during the growth phase of a technology.

I conclude that predictive ability of EIS and the effect of subsequent treatment are the most important factors to uptake at this stage of product development for the following reasons: 1. they were named as the most important factors in the final question. 2. Analysis of the language used in the interviews showed that they were considered most important. 3. Development of a strong evidence base precedes many other activities targeting the wide implementation of EIS.

Research evidence has the potential to affect diffusion in two ways: through reduction in uncertainty and through the produced study results. If a subsequent study showed that the predictive ability was worse than the current estimate, this would presumably result in lower uptake. Both effects should ideally be examined in subsequent studies. In the remainder of this thesis I focus on the reduction of uncertainty. I will also illustrate both effects, of reduction in uncertainty and study results, on diffusion and subsequent EVII analysis in a stylised example.

Figure 5.6 EIS lifecycle and possible relevance of implementation strategies



5.4 Discussion

5.4.1 Findings

In this chapter, I have provided an overview of relevant uptake factors for EIS and identified potentially effective implementation strategies at this stage of product development which were: research evidence on a) the predictive ability of EIS and b) treatment effect of progesterone compared with no treatment when EIS screening was performed. These findings are useful to product and study developers and policy-makers alike. The most important uptake factors are also relevant to health economists who may want to take the opportunity of assessing the value of implementation measures in expected value of implementation analyses.

5.4.2 Relevance

Although it is in the nature of a qualitative study to be context-specific instead of generalisable, these results highlight the potential importance of high quality research evidence to the adoption process, especially at an early stage of product development. This has implications for expected value of information analyses in which the assumption of independence of implementation and information was typically made (Fenwick et al., 2008). This assumption may still be satisfied if the evidence base is mature enough to not affect uptake anymore. This, however, raises the question of why EVI analysis should be performed in such a case: surely, the motivation of conducting further research is to ultimately affect implementation, whether in the shape of increasing implementation or encouraging the use of alternatives.

This study also adds further insight into the recent investigations on adoption barriers in the NHS. Whilst some of the identified themes matched the barriers and facilitators to adoption identified by the Department of Health (2011), it became clear that that report focused on a later stage of the product lifecycle and that our study adds some insight into uptake factors at an earlier stage, albeit with limited generalisability.

5.4.3 Strengths, limitations and further research

Qualitative methods appeared useful in identifying technology-specific uptake factors as they provide the freedom required to explore certain areas and the depth needed for exploring the detail. Strengths lay in the ability of eliciting the respondents' thoughts, associations and justifications without limiting them in their responses. Although difficulty in calling some factors to mind were observed, interviewees mentioned a broad range of different factors and theoretical saturation was achieved relatively fast, with all themes and sub-themes having emerged at the sixth interview.

It was not possible to identify the uptake determinant that can be translated into the most effective implementation strategies from the qualitative findings. A quantitative assessment of the effectiveness and costs associated with the different strategies is necessary for this. But with the aim being to select potentially effective implementation strategies for that quantitative assessment, I had to rely on some selection criteria that used the respondents' opinions and language as well as the lifecycle stage the technology was at.

Alternatives to these selection criteria would be performing a focus group to achieve consensus on the most effective implementation strategies or a survey in which all the different uptake factors could be rated in terms of their potential impact. I opted against a focus group or a survey because I thought that the gain in knowledge would have been minimal. This is because the potential impact of uptake determinants is context-dependent; for instance, it is dependent on the lifecycle stage that the technology is at. Furthermore, it would not have contributed anything to the purpose of this study, as the 'wrong' choice of uptake determinant would not impair the quality of the elicitation study, just yield the effect of a strategy that may not be the most effective. But even a focus group or a survey would not necessarily have given me the most effective, let alone, the most cost-effective implementation strategy, as a quantitative assessment of the potential effectiveness and costs of the implementation strategies would be needed for that.

Framework appeared to be an effective way of analysing and illustrating the data. It helped with getting an overview as well as a deep understanding of the themes and sub-themes. The subsequent analysis of effective uptake factors according to the respondents' statements and use of language helped identify potentially effective implementation strategies. Being a qualitative study, there is a certain degree of subjectivity in the data analysis. I did my utmost to work in a transparent way and document every step of the analysis in order to avoid my own potential bias to enter the analysis.

There is a possibility that not all factors relevant to uptake have been covered given that the sample size was relatively small and that the sampling frame was limited to two trusts. One respondent in the elicitation study expressed the view that some smaller units may be less likely to adopt innovation and although both University Teaching Hospital and District General Hospital types were covered, the included trusts were both large ones. However, the benefit of performing further interviews at different hospitals was considered to be relatively small compared to the time and effort required. This is because it was not particularly likely that smaller trusts would be driven by different factors. Instead it could be expected that some of the included factors would have a stronger influence on smaller units. Also, assuming that other themes did emerge, it is unlikely that these would change the overall result of this study. To verify this it would, however, be interesting to focus on smaller trusts in another study. The inclusion of more business managers may also be worthwhile to reflect their experience in different trusts.

A limitation was noted with the interview guide: when asking the question on external factors influencing uptake most respondents did not understand the relevance of this question and asked for more explanation. In an attempt to explain, the example of guidance or policies was provided. This type of leading question could prove problematic and therefore interviewees' responses were carefully examined. The result of this showed that respondents were genuinely convinced that guidance would play a major role and for most participants, other external factors then came to mind as well, such as the influence of media. In future studies, it would be worthwhile thinking about a better way of explaining external factors that avoids prompting, such as the description of influences that come from outside the local organisation, for example from authorities, other stakeholders and even other jurisdictions.

Scope for further research lies in quantifying the effect of the identified factors on uptake through elicitation of expert opinion.

5.4.4 Conclusion

In conclusion, qualitative interviews with framework analysis and the use of selection criteria can help identify those factors that influence uptake of individual health technologies and that may be targets of effective implementation strategies that can subsequently be evaluated in the context of economic evaluation.

CHAPTER 6. PREDICTING DIFFUSION OF ELECTRICAL IMPEDANCE SPECTROSCOPY

6.1 Introduction

One objective of this thesis was to identify a prediction method for diffusion that (a) could be used prior to technology introduction; (b) is relevant to the health technology context; and (c) yields probabilistic diffusion estimates. Literature reviews in Chapter 3 identified an expert judgement method based on a formal diffusion model to be the most appropriate method for this task in the health technology setting. In Chapter 3, I also showed that, so far, no such method has been described in the literature. Furthermore the Bass model of diffusion was identified as one such formal diffusion model, and I proposed the Satoh adaptation of it to address limitations associated with the Bass model.

In Chapter 5, the most relevant factors to the uptake of EIS were identified. When examining these factors for their ability to be turned into implementation strategies, I found that two different research studies, one on the predictive ability of EIS and one on the treatment effect of progesterone treatment after EIS screening, were important implementation strategies at this stage of technology life. In order to evaluate further research as an implementation strategy, counterfactual estimates of diffusion with and without the effect of further research are needed.

In this chapter, I wish to develop an elicitation of expert beliefs method that uses the Bass model of new product growth and the Satoh application of it, and to apply this method in the EIS for PTB screening case study. I predict natural diffusion and its uncertainty, as well as the effect of further research, in terms of its reduction of uncertainty, on future diffusion. Results of this study are used in health economic analyses that are presented in Chapters 7 and 8. The place of this chapter within the thesis is illustrated in Figure 1.2 in Chapter 1.

This chapter is structured as follows: Section 6.2 describes the methods used to develop the quantities to be elicited and to perform the elicitation of experts' beliefs exercise. In Section 6.3, I present the results of the elicitation study, including the individual and pooled estimates of diffusion and a comparison of the Satoh and the classic Bass model approaches in producing usable and realistic diffusion curves, as well as a reflection on the exercise. I conclude with a discussion and conclusion in Section 6.6.

6.2 Methods

6.2.1 Elicitation of expert beliefs

Elicitation is a widely-used method to obtain quantitative expert judgements reflecting uncertainty (O'Hagan et al., 2006). It describes '*the process of capturing expert knowledge about one or more uncertain quantities in the form of a probability distribution'* (O'Hagan and Oakley, 2010). Elicitation of expert opinions is inherently a Bayesian approach: it is often used where there are no frequentist observations about a research question of interest. Although elicited evidence is viewed as having the potential to inform health economic analyses whilst realistically representing uncertainty (Stevens and O'Hagan, 2002), its use in health technology assessment (HTA) is not yet widespread. However, its use appears to be on the increase (Sullivan and Payne, 2011, Soares et al., 2011, Leal et al., 2007).

6.2.2 What to elicit: re-parameterising the Bass model of diffusion

In this section, I discuss different options and considerations for eliciting information to populate the Bass model. I noted in Chapter 3 that it is generally perceived as difficult to invert observable quantities into Bass model parameters. Observable quantities that can be elicited from experts need to fulfil certain criteria in order to minimise bias: for instance, it is important that the summaries chosen for elicitation can be grasped intuitively (O'Hagan et al., 2006, Johnson et al., 2010). Care should also be taken not to confuse experts by eliciting too many different types of summaries, for example, eliciting a mix of absolute numbers, proportions and odds ratios (Soares et al., 2011).

Whilst one of the advantages of the Bass model over other models of diffusion was its behavioural parameterisation where parameters had an intuitive purpose, this does not necessarily mean that these parameters are straightforward to elicit. The Bass model requires knowledge of three parameters: 1. the maximum number of adoptions that can be achieved for EIS based on the current state of evidence, in the following called *m*. 2. The so-called coefficient of external influence, or the growth in uptake that occurs without any peer influence taking place, *p*. 3. The coefficient of internal influence, which reflects peer influence in the speed of adoption, *q*. I deemed quantity *m* to be straightforward as I expected experts to have beliefs about it. Elicitation of parameters *p* and *q*, however, would not be straightforward given that there is no easy way of quantifying peer influence or external pressures. I therefore thought these to be infeasible to elicit directly.

Instead, I developed an approach to inverting observable quantities to yield the parameters *p* and *q*. Two observable quantities were chosen to achieve that: (a) the point of inflection of the s-shaped curve t', i.e. the number of years at which the number of adoptions would start to decline, and (b) the initial number of adoptions in the first year after the technology was launched N_1. To invert these quantities to yield parameters *p* and *q*, I used an approximation method that entailed minimising the sum of squared error term. The squared error terms referred to the distance between the tested diffusion curve and *N1*, *t'* and *m*, respectively. The tested diffusion curves were obtained by using random parameter values for *m*, *p* and *q* to begin with. The curves were generated using two different versions of the Bass model, which I call the classic Bass model and the Satoh approach in the following. The classic approach was simply using the classic Bass model as presented in Chapter 4.4.1 to estimate per period adoptions. The Satoh approach used equations shown in Chapter 4.4.2 to estimate per period adoptions. It is worth noting that the classic approach required separate estimation of the number of adoptions for the first period using Equation (4.22) in Chapter 4. The Satoh approach avoided estimating first period purchases separately.

The approximation method by minimisation of sum of squared error terms, given *m*, N1 and *t'* was performed in Excel Solver. This entailed calculating the error terms between the elicited information and the Solver trial run diffusion curve for the three elicited quantities, and minimising the sum of the squared error terms. Within Solver, constraints were put on the parameters *p* and *q* as Excel was overwhelmed with the number of possible solutions. The constraints were set as follows: $0.000001 < p < 0.2$ and $0.000001 < q < 2.5$. These ranges were chosen such that they included parameter value ranges used in another study by Meade and Islam (2006) that were presented in Chapter 4. Ranges of likely parameter values of $0.00005 < p < 0.1$ and $0.01 < q < 0.99$ were reported in that study (Meade and Islam, 2006). While originally narrower ranges were chosen, they were widened to avoid truncating frequency plots when using the classic method. For both, the Satoh and the classic approaches, a number of 1,000 iterations of the solver algorithm was performed to obtain values for parameters *p* and *q*, by sampling from the distributions of the elicited parameters.

Other ways of inverting the obtained summaries were tried but did not yield better results. Initially, I tried to invert quantities algebraically, which would have been the exact and preferred way of obtaining the Bass model parameters. I attempted to obtain algebraic solutions for the above inversion problem using Bass model equations provided by Mahajan and Sharma (1986) and Sultan et al. (1990). These were presented in Chapter 4, Equations (4.23) and (4.22), respectively.

Solving both equations for *p* yielded, respectively:

$$p * e^{p*t'} = \frac{q}{e^{q*t'}} \tag{6.1}$$

$$p * e^{p} = \frac{\frac{1}{e^{q}} * \left(qN_1 + \frac{1}{q}\right)}{m - N_1} \tag{6.2}$$

Equations (6.1) and (6.2) do not have an algebraic solution. There may be other possible solutions but I did not find any more approaches of obtaining the Bass model parameters from observable quantities through algebraic calculations. While superior to an estimation approach in terms of its precision, an algebraic solution hence proved infeasible.

Another estimation approach using approximation was offered by Mahajan and Sharma (1986) who worked out *p* and *q* values using the following quantities: per period number of purchases at peak (*n'*) and cumulative number of purchases at peak (*N'*). For the Mahajan and Sharma (1986) approach to work, the cumulative number of purchases at peak was required to be smaller than $\frac{1}{2}m$ (half of total attainable adoptions), that is, $N' < \frac{1}{2}m$. This would result in diffusion curves that can only be asymmetrical in one way. There was no evidence to support this assumption, and I therefore discarded this option.

6.2.3 Elicited quantities

In summary, I chose three quantities that best informed the Bass model of diffusion based on current knowledge, and that matched the criteria of summaries suitable for elicitation. These quantities were: the maximum number of attainable adoptions (*m*), the number of years at which the number of adoptions would start to decline (*t'*) and the number of adoptions within the first year of technology launch (*N1*). These were elicited twice: 1. for the counterfactual scenario on which the currently undergoing trial would produce similar results for the predictive ability as were obtained from the pilot study; 2. given that two additional studies were performed, one on the predictive ability of EIS and the other on the relative treatment effect of progesterone versus no treatment after screening with EIS.

Of further interest was the extent to which the different research studies would contribute to the increase in uptake. One way of eliciting this information would have been to elicit the three summaries that inform the Bass model another two times, for each of the two research studies separately. With the time constraints imposed by participants, this was not feasible. I instead elicited the proportion of the increase in diffusion with both research

studies that was contributed by the progesterone treatment effect trial, assuming independence of the effects of both trials on diffusion.

As was mentioned in Chapter 1, another aspect to uptake that is not covered in this thesis is the extent to which a new technology is used once it is purchased. To obtain an idea of this utilisation in practice, I also elicited the proportion of patients being offered EIS once it is available at a hospital. The developers of EIS are currently studying the patient acceptability of EIS within the ongoing trial and I therefore did not deem it necessary to elicit this last aspect to technology uptake.

Below is a summary of all elicited quantities:

- the potential number of attainable adoptions (elicited twice, once given baseline evidence and once more given that both research studies have concluded) (m).
- the initial number of adoptions in the first year within product launch (elicited twice, once given baseline evidence and one more time given that both research studies have concluded) ($N1$).
- the number of years after which the number of adoptions starts to decline (elicited twice, once given baseline evidence and one more time given that both research studies have concluded) (t').
- the proportion of the increase in diffusion with both research studies that was contributed by the progesterone trial (compared with the study on the predictive ability of EIS) (κ).
- the proportion of patients being offered EIS once it is available at a hospital (π).

6.2.4 Individual elicitation meetings using SHELF

When a number of experts are invited for an elicitation exercise, there are two options of synthesising the elicited evidence. One option is to hold a group session and try to reach consensus within the meeting, a process that is called behavioural aggregation (O'Hagan et al., 2006) and has received increased attention (Sullivan and Payne, 2011). The other option includes individual elicitation sessions, followed by a type of mathematical aggregation of opinions. It is important to note that behavioural aggregation, that is finding consensus among experts, and mathematical aggregation methods can lead to different results (Soares et al., 2011). Advantages of a group session include the opportunity of letting experts reach a consensus, rather than choosing a method of synthesising their probability distributions

(O'Hagan et al., 2006). A drawback of the behavioural method is that it tends to produce over-confident judgements (Soares et al., 2011).

Without a consensus on the preferred method of aggregation, I conducted individual elicitation sessions which required mathematical aggregation of expert opinions. One reason for this choice was that it proved impossible to congregate experts at one place at the same time due to the participants' availability. Another reason was that a group session required not only individual elicitation exercises but also finding a consensus through additional elicitation sessions after discussion, all of which would have caused the group session to be significantly longer than each individual session. To maintain a group session within a reasonable time that participants were agreeable to (two hours), this would have required cutting down the number of quantities to elicit.

Previously, there were complaints about the lack of tools that facilitate elicitation methods (Leal et al., 2007). This research gap has now been addressed and there is a variety of tools and frameworks that can readily be used for elicitation of experts' beliefs (Leal et al., 2007, O'Hagan and Oakley, 2010). It was previously found that the use of a standardised script can help in establishing questions clearly (Johnson et al., 2010). I used the Sheffield Elicitation Framework (SHELF) in this elicitation study. SHELF is a formal procedure for elicitation of expert opinions (O'Hagan and Oakley, 2010), in particular focusing on group elicitation and finding consensus between experts. The decision to perform individual meetings rather than group meetings was made when some of the preparation was underway and therefore SHELF was kept as the package of choice. SHELF is a very straightforward package to use and although its greatest advantage lies in group elicitation, it performs well in individual elicitation and adds value through standardised guidance, processes and tools. SHELF provides tools that guide the facilitator through the process drawing on the insights of O'Hagan et al. (2006) but also on more recent practical experience (O'Hagan and Oakley, 2010). Elicitation is conducted in an open and well-structured way, which is regarded as being in accordance with best practice in the field (O'Hagan and Oakley, 2010). Supporting documents are provided, such as a pre-session pro forma, briefing and elicitation records and a bespoke R software package.

6.2.5 Population and sample

The population of interest were experts in the field of obstetric technologies. With this definition, it was mainly consultants in obstetrics that were included in the sampling frame,

although there was a chance that business managers might contribute another perspective on the issue of technology uptake. In the qualitative study in Chapter 5, I found that the key people in the decision-making process for the adoption of new health technologies in obstetrics were consultants, making them the primary population of focus.

The problem of overconfidence can be avoided by having a sample size greater than one (Johnson et al., 2010). A good sample size for an elicitation exercise is said to be no more than five experts, with three experts being sufficient as long as they are chosen based on their expertise (O'Hagan and Oakley, 2010). Table 6.1 summarises the professional categories that should be invited as well as the numbers to be included.

Table 6.1 Elicitation sample frame

Category to be included	**Number of individuals**
Obstetricians (UTH)	At least 1
Obstetricians (DGH)	At least 1
Business manager (UTH or DGH)	0-1
Sum	At least 2

I contacted participants in the qualitative study who had expressed their agreement with being contacted for a follow-up study through the informed consent form. Three experts (two obstetricians from the UTH and one from the DGH) agreed to participate. Their availability was assessed via email communication as there was reluctance to use a doodle poll. When it turned out that a group meeting was not possible, places and times were booked for each individual expert. All meetings were held at the participants' work places to ensure convenience.

To make sure that the sample of experts represented knowledge of technology uptake in obstetrics, experts were chosen based on the following characteristics that experts for an elicitation study should demonstrate, developed by O'Hagan et al. (2006):

- Tangible level of expertise: participants were selected if they had professional training and experience as an obstetrician and had experience in PTB screening.

Additionally, business managers of obstetric units were also considered for their experience with the purchasing of new equipment.

- Understanding of the general problem area: experts should have an interest in and an understanding of the problem of uptake of health technologies in obstetrics and the factors that could influence it. I made sure this was the case by approaching participants in the qualitative study who had expressed their interest in participating in a follow-up study on this topic.
- Reputation: experts should all be recognised as professionals in their area and should be deserving of other experts' respect. I assumed that this was a given, with all experts having been medically trained and none of them being a junior member of staff. Whether they had published in the area of preterm birth was not seen as important as this study was not of strictly medical nature.
- Lack of an economic or personal stake in research findings: if experts did have a stake in the findings of this research this would have to be laid open. Participants from the Sheffield Teaching Hospital could be suspected of having a stake in the study outcome as EIS was developed at their hospital. However, none of the invited consultants were directly involved with its development, or research surrounding it. Experts were also asked to highlight any stake or personal bias in the pre-elicitation pro forma and the issue was discussed at the introductory part of the elicitation. If potential biases had been identified, experts would have been encouraged to view the problem from different point of views or take other experiences into account. However, no biases were identified in any of the participants.
- All available expertise should be covered: the available expertise in this study related to professional categories and hospital type. Consultants of obstetrics were seen as the most relevant group with the ability to assess future uptake. I thought that different hospital types may influence the likelihood of new technology uptake and therefore made sure to have both, University Teaching Hospitals and District General Hospitals, represented by a consultant.

A pilot was conducted with one volunteer clinician. Garthwaite et al. (2005) flagged up the usefulness of a test run to obtain practice with the protocol. The pilot turned out to be helpful in practising the exercise because it featured a considerable number of quantities to be elicited and required a large amount of background information to be provided. The pilot also ensured effectiveness of the questions asked, gave me confidence with using the software and helped in assessing the time that the meetings would take to ensure that participants were adequately informed. The pilot was undertaken in November 2013 and three experts were interviewed in individual meetings between November 2013 and January

2014. Two weeks separated the pilot and the first meeting in order to ensure that all potential problems could be addressed. The elicitation meetings took between one and a half and two hours.

6.2.6 Ethical considerations

As with the qualitative study, the elicitation study had obtained ethics approval by the University of Sheffield Research Ethics Committee (UREC), as well as NHS R&D governance by the Sheffield Teaching Hospitals NHS Foundation Trust, the Rotherham NHS Foundation Trust and Barnsley Hospital NHS Foundation Trust. Confidentiality and anonymity were guaranteed. Written informed consent was obtained from all study participants after questions were addressed. The information sheet, informed consent form and elicitation briefing notes can be found in the appendix (Appendices D.1, D.5 and D.2, respectively). All information was sent well in advance of the meetings, at the recruitment stage.

6.2.7 Elicitation study design

The goal of an elicitation is to accurately represent an expert's knowledge (Garthwaite et al., 2005). As with all other research, transparency in every step of the research is important and can contribute to minimisation of bias in elicitation of expert opinions. There are recommended steps and proceedings provided by SHELF (O'Hagan and Oakley, 2010), which are informed by O'Hagan et al. (2006) that help ensure transparency and rigor in performing and reporting elicitation exercises. These have been adopted and adapted to the present study and their application in this elicitation exercise is described below.

6.2.7.1 Preparation

It is important that participants know what to expect and are prepared for the tasks required of them in the elicitation meeting. Preliminary briefing materials were thus sent in advance of the meetings. These included three documents (see Appendix D): 1. an information sheet detailing the purpose of the study, expectations of participants and ethical questions, amongst others (Appendix D.1); 2. the pre-session briefing notes which are a general description of elicitation exercises and are provided by SHELF (O'Hagan and Oakley, 2010) (Appendix D.2); and 3. the pre-session pro forma, also provided by SHELF

(O'Hagan and Oakley, 2010), that has the purpose of asking participants about the scope of their expertise and any potential conflicts of interest (Appendix D.3). Participants were asked to fill in the pre-session pro forma ahead of the meeting and return it to me. In the event of it not being returned, the form was filled in at the meeting itself. A fourth document was sent closer to the meeting: the background information sheet (Appendix D.4). It is described in more detail in the section on structuring below. At each meeting, participants were again provided with the information sheet and given time to read through it and raise and have answered any questions. They were then asked to give their consent to participate via the elicitation consent form (Appendix D.5).

6.2.7.2 Training

Training was provided to all participants to familiarise them with the use of probability distributions and improve the normative goodness of their estimates (Johnson et al., 2010). The training exercise was conducted at the start of each individual session and one quantity was elicited. Training quantities are typically quantities that can in principle be known (and are known by the facilitator) but are not likely to be known exactly by the experts. O'Hagan and Oakley (2010) recommended the use of a quantity that lay in the domain of the expert. I thus chose the number of pregnant women with a preterm delivery in England. This quantity was deliberately elicited in absolute terms, given that most of the quantities of interest were absolute numbers. Experts were then shown their calibration, that is the match of their estimates with the real data, and given feedback on their performance.

Offering training also provides an opportunity to discuss common heuristics, that is, strategies used for problem-solving that may induce bias. Such heuristics may be used by experts to assess probabilities. Hogarth (1987) as cited in Garthwaite et al. (2005) provided a more comprehensive overview of heuristics and biases and the more commonly used heuristics were described by O'Hagan et al. (2006), and are briefly summarised in Table 6.2.

Table 6.2 Commonly used heuristics

Availability heuristic	Describes how easily instances come to mind, a typical error being the over-estimation of probabilities of events that have, for instance, been covered in the news
Representativeness heuristic	Refers to the similarity between instances and expectations of those, common errors being:
	1. Conjunction fallacy: where people tend to think that $\mathrm{P(A)} < P(A \cap \mathrm{B})$. To give an example, people may think that the probability that someone is a smoker and delivers their baby prematurely is higher than the probability that someone delivers their baby prematurely.
	2. Base-rate (neglect) effect: ignoring the a priori probability of an event. For instance, when someone over-estimates the probability of a patient delivering their baby prematurely because of the patient being a smoker and ignore the a priori likelihood of preterm delivery.
	3. Small sample size inferences, also called hasty generalisation.
	4. Confusion of the inverse: $P(A\|B) \neq P(B\|A)$. A popular example is that of there being a high probability that hard drug users also use marijuana. That does not lead to the conclusion that there is the same high probability that marijuana users also use hard drugs.
	5. Insufficiently regressive predictions: people may change their predictions of an event based on new knowledge available but disregarding other factors that influence that event as well, thus over-estimating the change in the event.
Anchor-and-adjustment heuristic	Describes anchoring a decision at an initial starting value and just make small adjustments, with common errors including
	1. Insufficient adjusting
	2. Pruning effect: where having a certain number of categories implies an anchor at $\frac{1}{\#\ categories}$ per category.
Affect heuristic	Describes when judgement is influenced by one's mood or risk adversity. For example, high risk procedures may be judged as being less beneficial than low risk ones.

The use of these heuristics can lead to the introduction of bias in the elicited probability distributions. One way of avoiding the use of heuristics is making the experts aware of them. For a lack of time, not all of these heuristics could be explained and discussed but I was aware of them and tried to avoid their use by probing for rationales at the elicitation of every quantity. One of these heuristics seemed particularly likely to affect this exercise: insufficiently regressive predictions in the elicitation of the effect an implementation strategy would have on the participants' uptake estimates. Participants may have been susceptible to focus on the effect of the implementation strategy and ascribe too big an impact to it by disregarding other factors at play. Participants were thus asked not to disregard all the other factors affecting uptake that they had identified previously when making their judgements.

6.2.7.3 Structuring: providing relevant background information

Relevant information and knowledge were brought to mind at this stage. The background information document (Appendix D.4) was thus presented to the participants and any questions they might have clarified. Background information included presentation of the evidence on the predictive ability of EIS obtained from the pilot, evidence on the progesterone treatment effect, a description of the device, an assumption that capital and disposable costs would not be prohibitively expensive and a plausible range of number of devices needed per hospital.

It is worth noting that the uncertainty the experts described could be of aleatory nature, that is, it could be caused by inherent variabilities or randomness in the system, or it could be epistemic, that is, due to imperfect knowledge and can therefore be eliminated by specifying more conditions (O'Hagan and Oakley, 2004). Uncertainty in this context is more likely to be of epistemic nature: if the state of more independent variables were known, experts would likely be able to give a more accurate estimate of the probabilities they were asked for. Providing the background information described in this section was one way of specifying conditions that reduced this epistemic uncertainty. One of these specified conditions was that study results would be similar to existing evidence. This was to capture the effects a reduction in uncertainty rather than the effects of different study results but the caveat is that possible difference in study results are not explicitly considered in the diffusion estimates.

To cover background on EIS and its evidence, I presented positive and negative predictive values and likelihood ratios that were available at that point in time, and the size

of the study that was being conducted at that point. Participants were asked to assume that the study currently underway would find similar results to the ones that were available from the pilot study while taking uncertainty about the mean estimates into account. The findings from the qualitative study were also briefly discussed, with a focus on further research influencing the diffusion of EIS. In accordance with the qualitative study findings, further research was broken down into evidence on the predictive ability of EIS and the relative treatment effect of progesterone compared to no treatment after using EIS screening. It was then agreed that both research studies would have a similar study design and sample size to existing or ongoing research studies that were cited in the background information sheet (Appendix D.4). Participants were informed of effectiveness estimates for progesterone when administered after testing positive with cervical length scans, bearing in mind that these estimates may differ after screening with EIS.

As the price of the technology was not known at the time, it was assumed not to be prohibitively high but participants were asked to keep some uncertainty regarding that in mind, too. Information was provided on the number of obstetric units in England and the mean and range of their sizes in terms of patient numbers in order to enable experts to make judgements on the number of devices that could be purchased at maximum.

Furthermore, evidence for the general shape of diffusion curves was presented in order to show respondents possible outputs of this study. Diffusion curves informed by a meta-analysis as well as data for two other technologies were shown. These different diffusion curves were deliberately chosen to illustrate differences in the speed of diffusion. The main benefit in presenting these curves lay in visualising the elicited quantities and their potential effect on diffusion curves. There may have been a possibility that showing these curves made respondents more prone to using an anchor-and-adjustment heuristic and they were therefore encouraged to avoid using the presented diffusion curves as indicative of diffusion of EIS. In addition to that, when it came to eliciting the relevant quantities, experts were encouraged to think about their own experiences in this therapeutic area as well as about the background information on the number of obstetric units in the country (Appendix D.4).

6.2.7.4 Structuring: the elicitation

The first task was to elicit the ‘baseline’ diffusion; that is, diffusion of EIS conditional on 1. the ongoing trial delivering similar results to the pilot study and 2. the relative progesterone treatment effect being similar after EIS screening to after cervical length scans. The second task was to elicit diffusion conditional on the additional two studies that were

identified to have an impact on uptake having concluded. The third task was to elicit the contribution of the progesterone study to the change in diffusion. For this, it was assumed that the research studies were independent in their effect on diffusion (that is, research on the treatment effect would have a certain effect on diffusion regardless of what other research was conducted, and vice versa). The last task was to elicit the extent to which EIS was likely to be used by individuals after it had been purchased by the trust.

All elicited quantities (*m, N1, t'*) were assumed to be independent. When eliciting multiple quantities from one expert, some thought has to be given to possible dependence between responses (Soares et al., 2011). There was some indication that quantities might be correlated, for example the number of attainable adoptions and the number of adoptions in the first year, but multivariate elicitation would have significantly increased the complexity of the exercise (O'Hagan et al., 2006) and experts may not have been able to provide the level of detail needed for eliciting joint probabilities (Soares et al., 2011) or other quantities required to assess correlation (that are described by Daneshkhah and Oakley (2010)). SHELF does not recommend elicitation of multivariate problems for the same reasons (O'Hagan and Oakley, 2010). I therefore excluded parameter inter-dependence from this analysis and elicited only the marginal distributions. This was discussed with the participants.

6.2.7.5 Probability distributions

A parametric method of fitting probability distributions was chosen for this study. Different methods of fitting probability distributions were discussed by Garthwaite et al. (2005). Parametric methods impose the structure of a member of a family of distributions by choosing appropriate hyperparameter values that represent the expert's opinion (Garthwaite et al., 2005). Non-parametric methods of fitting distributions may pose the advantage of avoiding restrictive assumptions and forcing the elicited probability distribution to fit a parametric family (Oakley and O'Hagan, 2007) but they are more complex and may not actually yield a considerable benefit as imprecisions often have very little effect (Garthwaite et al., 2005). In this study, the obtained probability distributions were presented to the experts for them to confirm that they were consistent with their beliefs. I was satisfied that any slight imprecisions were not likely to significantly affect the cost-effectiveness model and expected value of implementation and information analysis that these data would feed into because imprecisions would have a small effect on cost-effectiveness and value of implementation. In summary, parametric methods were chosen for three reasons: first, to avoid the added complexity of non-parametric fitting; second, because the benefit of

obtaining a slightly more accurate distribution was questionable; and third, because parametric distribution fitting was embedded in SHELF and the resulting distributions were thus readily available. The SHELF software allows choosing the best fitting distributions from normal, log-normal, beta, scaled beta and gamma distributions. Gamma distributions were deemed appropriate for all parameters except for κ and π, for which beta distributions were used.

The variable interval method was chosen as opposed to the fixed interval method (Garthwaite et al., 2005, Oakley et al., 2010). The former refers to experts making quantile judgements and the latter to the expert making judgements on the probability of the uncertain parameter θ falling into pre-specified intervals. Fixing pre-specified intervals constrains results compared to variable interval methods but fixed interval methods may be easier to complete. Which method performs best is debated and both Oakley et al. (2010) and Garthwaite et al. (2005) suggest that there is no conclusive evidence.

SHELF provides different protocols for different methods, among them variable interval methods such as tertile and quartile methods and a fixed interval method that is called the roulette or histogram method (Soares et al., 2011, Bojke et al., 2010, Oakley et al., 2010). The roulette or histogram method divides the range into intervals (bins) and asks experts for the probabilities of those intervals. This method provides the advantage of being very visual and potentially easier to understand than variable interval methods. However, the process of assigning probabilities for many intervals and each quantity could become quite repetitive. In terms of variable interval methods where experts are asked to shift the interval to yield a certain probability, tertiles have been shown to yield better calibrated probability distributions than quartiles (O'Hagan et al., 2006). This is potentially a result of over-confidence presenting a greater issue with the use of quartiles than with the use of tertiles (Garthwaite et al., 2005). Based on the greater flexibility and avoiding repetitive tasks and over-confidence, I chose to elicit tertiles.

6.2.7.6 Elicitation steps

For each elicited quantity, the first step entailed asking experts to define the lower and upper bound such that it was extremely unlikely to impossible that the quantity to be elicited fell outside this range (this was specified to be close to the 0.1^{st} and 99.9^{th} percentiles). Experts were asked to think carefully about that range so as not to choose a range too wide or narrow. I then asked for the median and for the upper and lower tertiles. All of these values were written down in the elicitation record 2 (D.7), with experts' rationales added

wherever appropriate. A distribution could then be fitted to the experts' assessments. At this stage, accuracy of the assessment was improved by verifying that experts were happy with their elicited distribution. This was done through presenting their probability density function and through confirming the implied values at the 10^{th} and 90^{th} percentiles. If experts were unhappy with these, the distribution was changed. After adjustments, the final distribution was recorded.

All these steps were undertaken in all meetings and results were recorded in the elicitation record 1 and 2 forms (Appendices D.6 and D.7), provided by SHELF (O'Hagan and Oakley, 2010). The elicitation record 1 was filled in for each expert and contained a summary of the information available on the expert as well as on the summaries elicited. The elicitation record 2 was filled in for each expert and all of the quantities elicited from them. This served the purpose of recording the experts' beliefs as well as any qualitative explanations.

To enable reflection upon the exercise, I asked all the participants for brief feedback on the elicitation session at the end of the meeting. Together with my own observations on the meetings, this feedback was recorded in a post-elicitation record (Appendix D.8) to allow evaluation of expert elicitation as a method of obtaining diffusion estimates in this setting.

6.2.8 Data analysis

Once the elicitation meetings were conducted, several activities were undertaken to analyse the data. First, a closer examination of the resulting pooled diffusion curves was of greatest interest. Second, individual elicitation results were examined separately. Third, feedback and observations on the process of this study were evaluated.

6.2.8.1 Obtaining pooled estimates

Diffusion was presented in terms of number of adoptions in each period. I thought about representing uptake as a proportion of all desirable adoptions, however, it was not clear at what number of adoptions 100% was reached. For example, if it was assumed that, to make sure that all high risk patients in the country are covered, every obstetric unit would have to have at least two devices, then the desirable number of adoptions would be at least 462. However, some of the obstetric units are very small and may refer high risk patients to larger units. Some of the very large units may need a greater number of devices to ensure that one

is available for all patients at all times. I therefore opted against using proportional uptake going forward and focused on absolute numbers in the elicitation study. It is important to note that participants may have used their own assumptions on how many adoptions were desirable when providing their estimates on elicited quantities.

Participants' opinions on diffusion were aggregated mathematically. There are two methods of doing that: opinion pooling and Bayesian methods (Soares et al., 2011). Bayesian methods can be difficult to implement as they are based on the assumption that the aggregate distribution reflect the belief of a 'supra Bayesian' decision-maker (O'Hagan et al., 2006, Genest and Zidek, 1986). Opinion pooling, and linear pooling in particular, was more commonly used in elicitation (Cooke, 1991) and was found to perform well (Clemen and Winkler, 1999).

A linear opinion pool was chosen to be the most suitable type of mathematical aggregation. When using opinion pooling, there are two approaches: linear and logarithmic pooling. Logarithmic pooling offers an advantage when experts come from different backgrounds and it is expected that each of their experience adds to obtaining a more accurate estimate. Linear pooling on the other hand treats experts as being more fallible (O'Hagan et al., 2006) and thus reduces overconfidence (Johnson et al., 2010). The experts that contributed to this study had similar backgrounds and where dissimilarities occurred (for instance, in the length of time they had worked in their field) it was difficult to assess whether this might influence their calibration (none of the experts were very junior). The assumption that their ranges of knowledge would contribute to a better estimate did therefore not seem warranted. Hence, the more conservative option to use, that is the option that would result in greater uncertainty, was linear pooling.

The arithmetic mean was not weighted in this study's linear pool. A weighted arithmetic mean can be used where the expertise of some participants is clearly superior to those of others. This could not be attested through the preliminary questions asked in the pre-session pro forma (Appendix D.3). Another way of checking for participants' superiority in expertise is through their calibration in the training exercise (Soares et al., 2011). Soares et al. (2011) highlighted that it was not clear how to choose appropriate seed questions and that research on how to obtain weights to reflect experts' relative expertise was needed. In this particular elicitation, better calibration in the training exercise did not indicate better calibration with regards to future diffusion of EIS. With those difficulties of judging the expertise of each participant and the questionable benefit, I used an un-weighted arithmetic mean. For the linear pool, the arithmetic means of all elicited points of the distributions were taken. The latter included the lower and upper range, tertiles and the median.

6.2.8.2 Analysing the pooled diffusion curves

i. Comparing the performance of the classic and the Satoh approach

The Bass model parameters for the pooled diffusion curves that were estimated by inverting the probability distributions of elicited quantities using both the classic Bass and the Satoh approaches, as described above, were examined for the following:

- How the individual parameters *p* and *q* affected the diffusion curve was examined through holding *m* and *q* (*m* and *p*) fixed at their means and exploring different values for the remaining parameter. These different values were the mean, maximum and minimum values that resulted from the simulation using excel solver. The impact of those different values for *p* (*q*, respectively) were plotted in figures showing diffusion curves with both the classic and the Satoh models.
- Of particular interest were the parameters *p* and *q* as they were not elicited directly but approximated using solver. Their approximated distributions were plotted in density plots for both approaches.
- The potential correlation of *q* and *p* while *m* was fixed at its mean was explored using smooth contour plots for both approaches.
- In order to see whether *q* and *p* were correlated with *m*, scatter plots were created for both approaches.
- Five examples of likely *p* and *q* combinations from the highest density level in the smooth contour plots were plotted in cumulative diffusion graphs for both approaches.
- Thirty randomly sampled diffusion curves were plotted in graphs for both the classic Bass and the Satoh model.

ii. Examining the cumulative diffusion of EIS with and without research

Only the Satoh method was used for this part of the analysis as it was shown to perform better at obtaining realistic diffusion curves than the classic Bass model. The two curves for diffusion with and without further research were plotted in one graph. The proportional effect of the progesterone trial on diffusion when compared with the study on the predictive ability of EIS was applied to the difference of the adoptions in each period with and without research. The resulting diffusion curve was plotted on the same graph as well. One expert felt unable to give any estimate for this quantity and a weakly informative (flat) uniform distribution covering the range from 0 to 1 was thus used for the pooling. The mean diffusion estimates resulting from the pooled data and the solver algorithm applied to yield

parameters *p* and *q* were compared with the pooled elicited quantities. For this, the pooled mean values and standard deviations for each quantity were presented in a table and the pooled and individual probability distributions were plotted in probability density functions.

6.2.8.3 Examining individual diffusion data

Individual diffusion curves per participant were generated to examine how they differed amongst each other. For each elicitation, this required running 1,000 simulations of the solver algorithm for diffusion with and without further research to yield the likely space of diffusion curves. Qualitative explanations provided by the participants were analysed to shed light on how participants arrived at their estimates and how their rationales translated into different shapes of diffusion curves.

6.2.8.4 Reflection on the exercise

I considered how results of this study could be validated and assessed for any potential bias. It is not clear how validity, reliability and responsiveness can be measured in elicitations of probability distributions (Johnson et al., 2010) and I therefore attempted to ensure minimisation of bias through recommended measures, including my study design, the choice of experts, the type of elicited quantities and the use of a standardised elicitation tool (Johnson et al., 2010). Because elicitation of expert opinion can be a complex exercise and demanding on the participants who may not be used to thinking in terms of probability distributions, I also evaluated the use of this method by reflecting on it with the help of my own observations, feedback that was provided by the experts and elicitation results.

6.3 Results

6.3.1 Pooled distributions and performance of methods

6.3.1.1 Performance of the classic versus the Satoh approach

To enable an understanding of the effects of *p* and *q* values on the shape of the diffusion curve, the impact of different values of *q* and *p* in the classic Bass model are shown in Figure 6.1 and Figure 6.2. These values are derived from the solver algorithm applied to the

pooled elicited quantities. It appears that *q* has a more pronounced effect on the diffusion curve. While different values of *p* result in curves that mainly shift left or right, most significantly altering the number of adoptions in the earlier periods, different values of *q* can delay the process of diffusion to a very slow one (in the case of the minimum *q* it would take more than 50 years to reach the attainable number of adoptions) or even compromise the s-shape altogether, as seen in Figure 6.2.

The extreme and unrealistic shape of the diffusion curve caused by the maximum value of *q* stems from numerical problems caused by independent elicitation and the discrete nature of the continuous function. Parameters *p* and *q* were estimated by approximating to the number of initial adoptions and the time it takes to reach the maximum number of adoptions in one period. With the combination of high values of the former and low values of the latter, both *p* and *q* would be estimated at large values. The curve would grow very fast and exceed the maximum cumulative number of adoptions which then leads to negative adjustment in the following period. Of course, negative growth of cumulative adoptions is impossible, so this is an obvious limitation with the use of this model.

Figure 6.1 Effect of minimum, maximum and mean p on the shape of the classic diffusion curve

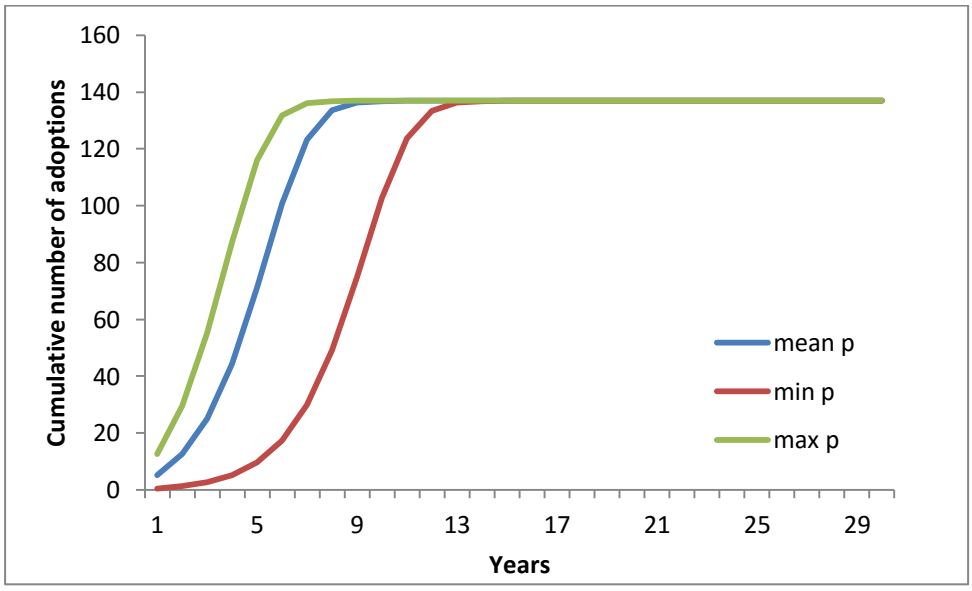
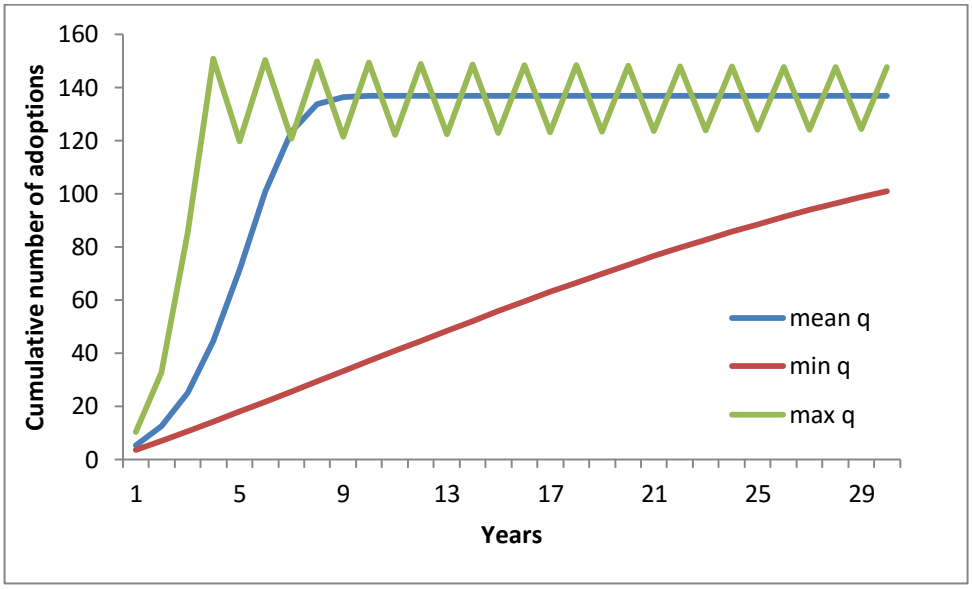


Figure 6.2 Effect of minimum, maximum and mean q on the shape of the classic diffusion curve



This limitation is a result of the univariate elicitation of a multivariate problem that results in implausible combinations of parameters as well as the fact that the Bass model is time-continuous but represents discrete data. The Satoh adaptation of it addresses these problems and the resulting diffusion curves with the simulated mean, minimum and maximum values for *p* and *q* are indeed s-shaped as plotted in Figure 6.3 and Figure 6.4.

Figure 6.3 Effect of minimum, maximum and mean p on the shape of the Satoh diffusion curve

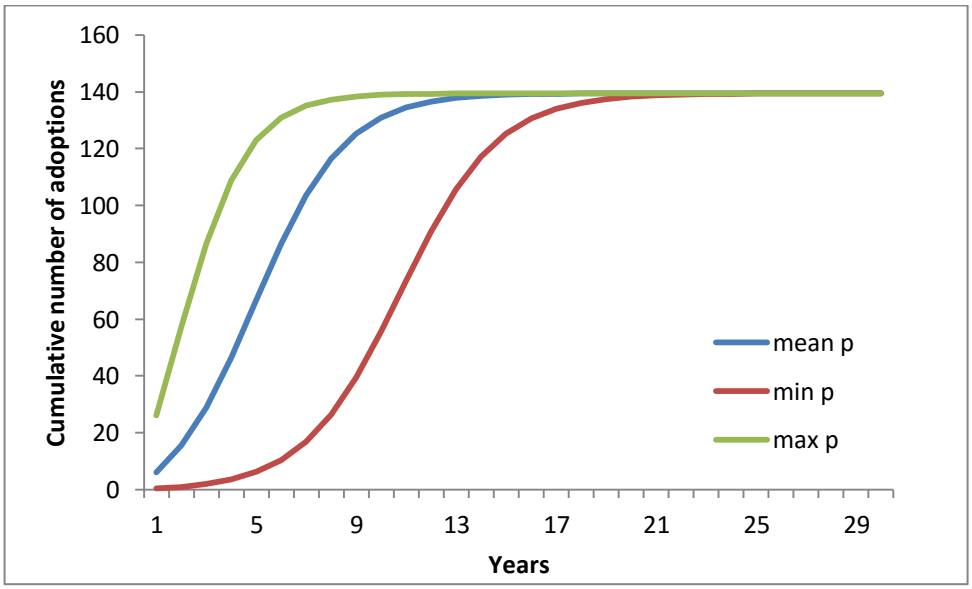
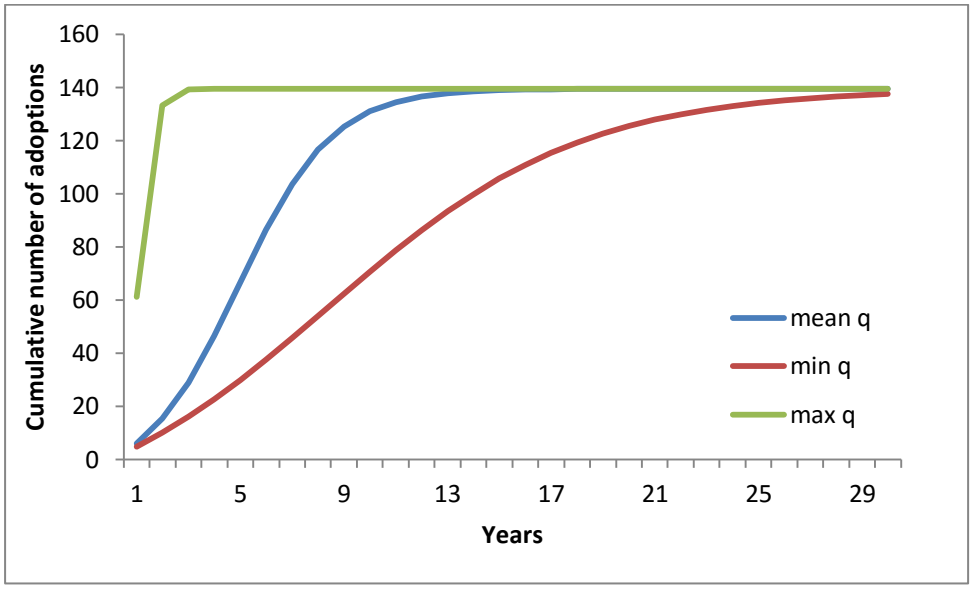


Figure 6.4 Effect of minimum, maximum and mean q on the shape of the Satoh diffusion curve



Whilst the Satoh approach does not affect the estimation of the quantity of attainable adoptions which was elicited directly, its obvious impact is that the resulting parameters *p* and *q* fall into different ranges. Parameter *q* does not assume values greater than one with the Satoh approach while the classic approach allowed it to assume values as large as two. Parameter *p* on the other hand, assumes larger values with the Satoh approach. This can be seen in the distributions of parameters *p* and *q* resulting from applying the solver algorithm to the elicited values that are plotted in Figure 6.5 and Figure 6.6. Being conditional distributions, *f(p|m)* and *f(q|m)* do not follow any standard parametric distribution, and only the density plots for the marginal distributions of *p* and *q* are shown.

Figure 6.5 Density plots of p and q for the classic Bass model

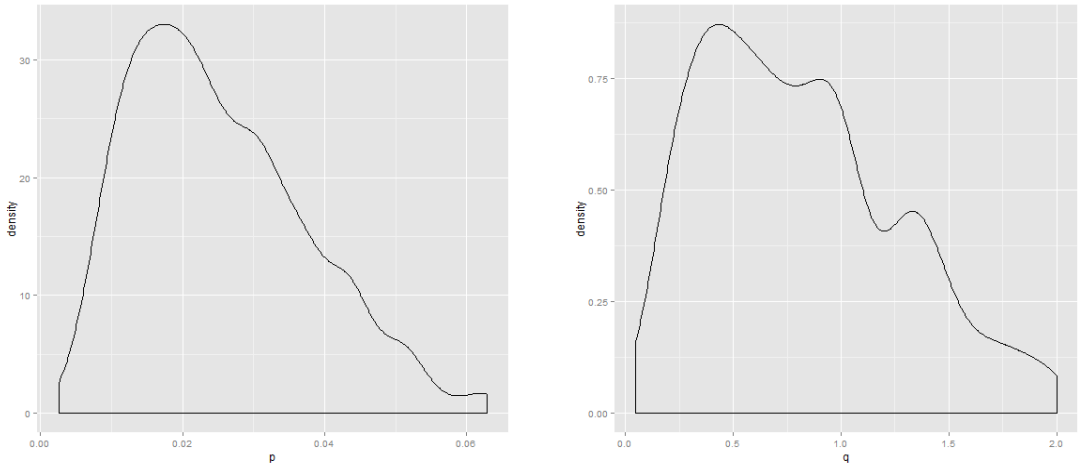
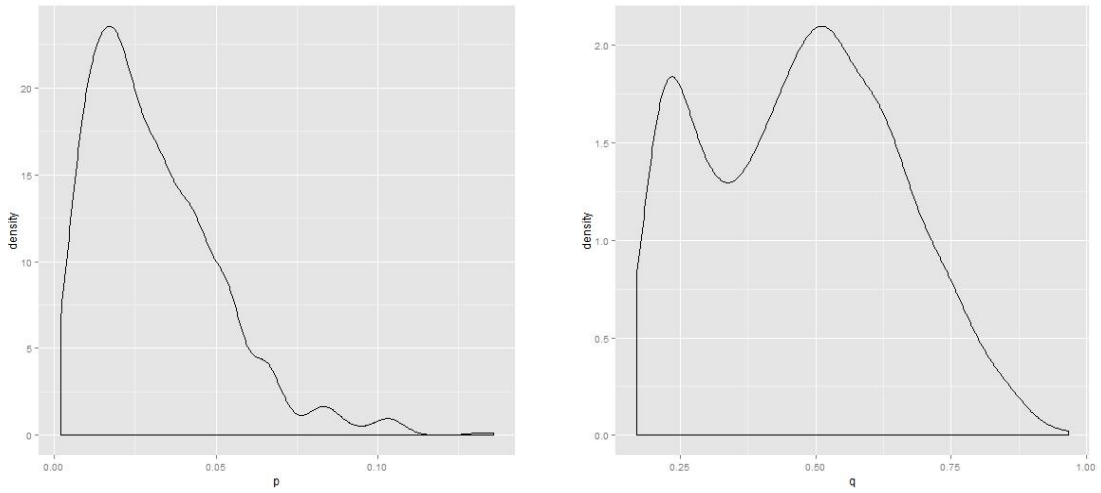
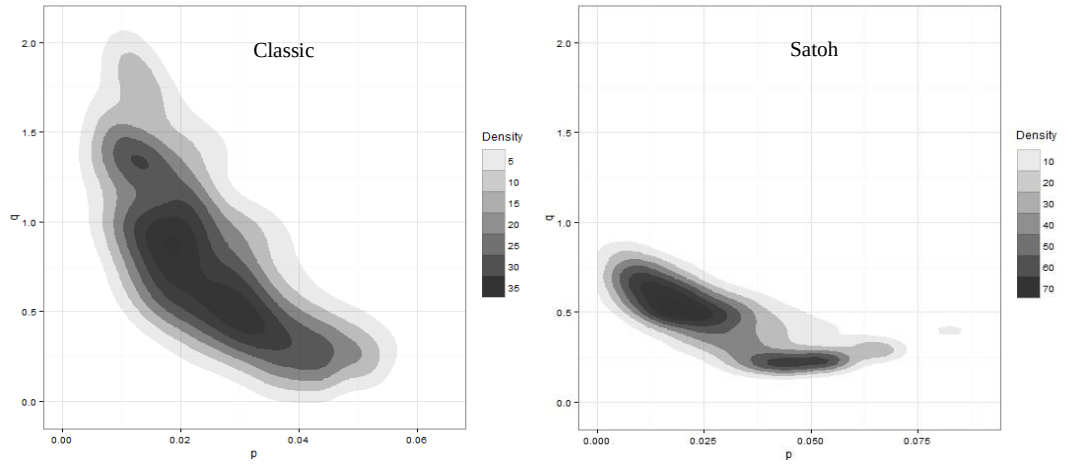


Figure 6.6 Density plots of p and q for the Satoh approach



Despite the elicitation being performed univariately, the implied *p* and *q* distributions are correlated. This is illustrated in Figure 6.7, which shows concentrations of *p* and *q* combinations in smooth contour plots for the classic and the Satoh models. There is a trade-off between parameters *p* and *q*. The larger *p* is, the more likely *q* is to be small and vice versa. The Satoh method results in *p* and *q* combinations that are more concentrated. One reason for this could be that the Satoh approach, while preventing unrealistic shapes of the diffusion curve, implicitly limits the value of *q*.

Figure 6.7 Smooth contour plots for the correlation of q and p with the classic and the Satoh approach



The negative correlation between *m* and *p* that is illustrated in scatter plots on the left hand sides of Figure 6.8 and Figure 6.9 is a reflection of the functional relationship between *p* and the number of adoptions in the first period N_1. Parameters *p* and *q* are functions of that number N_1 and the time to the point of inflection of the diffusion curve *t'*. The quantity N_1 was obviously positively correlated with *m*, as the larger the attainable number of adoptions, the larger should be the number of adoptions in the first period. Because of the assumption of independence of N_1, *t'* and *m*, N_1 would be distributed in the same way irrespective of any changes to *m*. So, if *m* adopted a smaller value, the resulting ratio of N_1/m would become larger and, because *p* reflects larger numbers in initial periods, *p* would therefore adopt larger values.

There is also a positive correlation between *m* and *q*, illustrated in Figure 6.8 and Figure 6.9 on the right. This is less pronounced than the negative correlation of *m* and *p*. If the correlation between *q* and *m* was indeed positive, this could be explained through the N_1/m ratio being smaller with larger values of *m*. Because *q* reflects a quick growth, this would result in *q* being larger to reach the attainable number of adoptions despite the small N_1 (or *p*).

Figure 6.8 Scatter plot showing the correlation of m and p, m and q for the classic approach

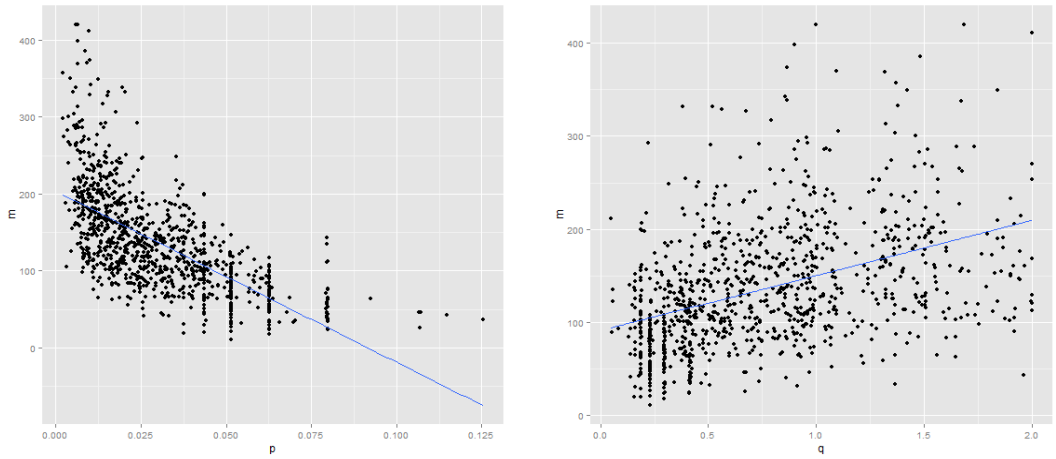
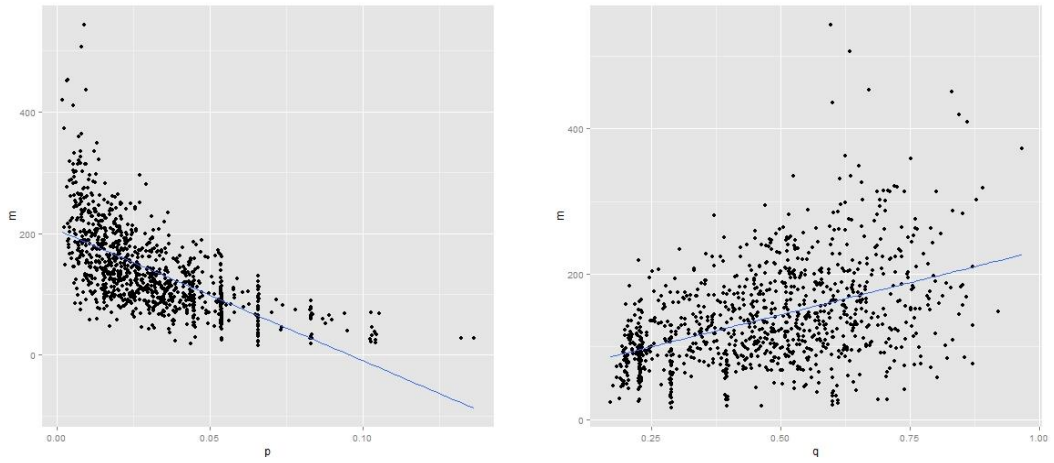


Figure 6.9 Scatter plot showing the correlation of m and p, m and q for the Satoh approach



In order to examine likely shapes of the EIS diffusion curve, cumulative uptake with five likely combinations of *p* and *q* is presented in Figure 6.10 and Figure 6.11 for both approaches. The s-shape of the diffusion curve is retained when choosing likely combinations of *p* and *q* in both methods. While small differences in the approaches can be observed through visual inspection, it is difficult to assess whether they matter or which method leads to curves that fit the experts' opinion better. Going forward, I used the Satoh adaptation of the Bass model, as it avoids the limitation of yielding unrealistic and unusable diffusion curves.

Figure 6.10 Five likely shapes of diffusion curves using the classic approach

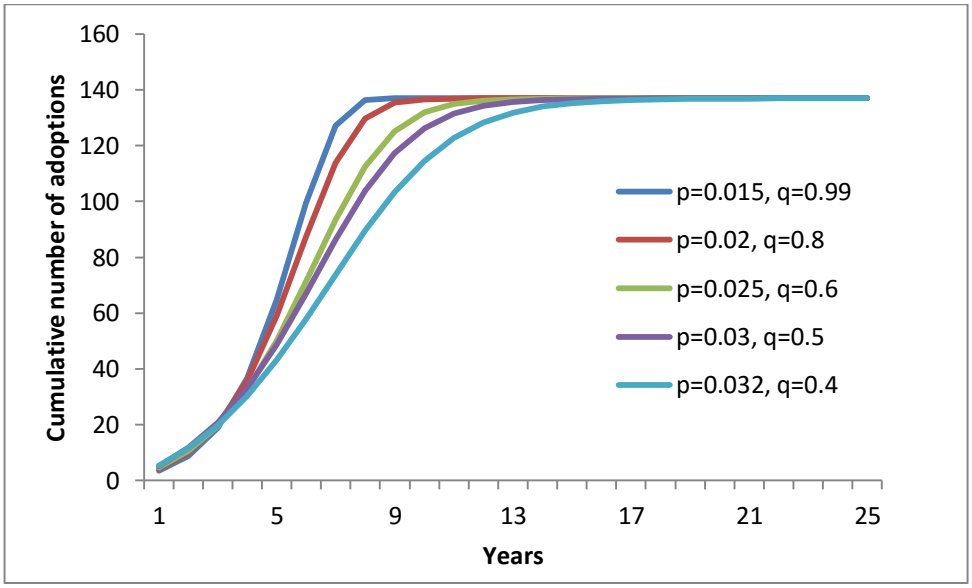
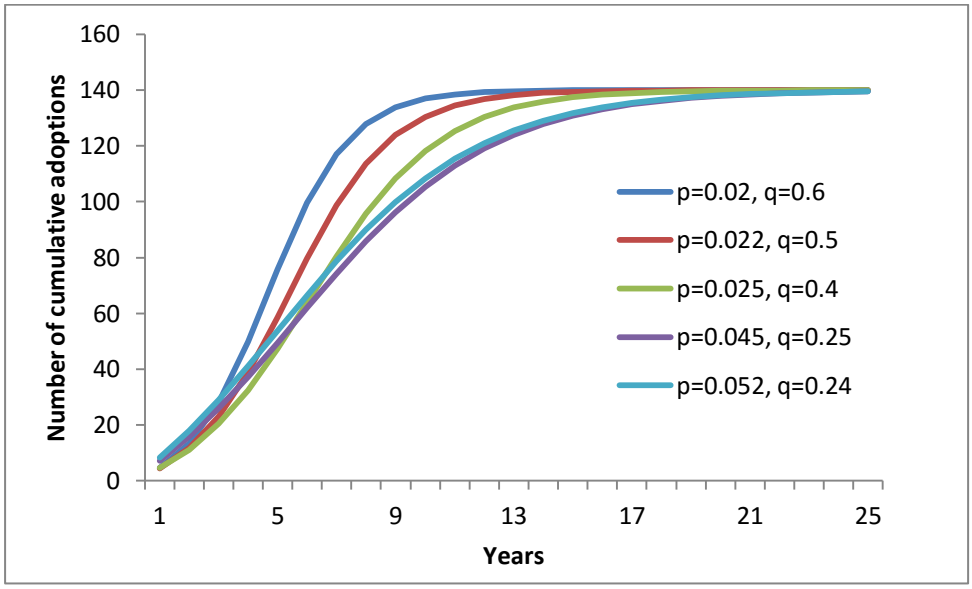
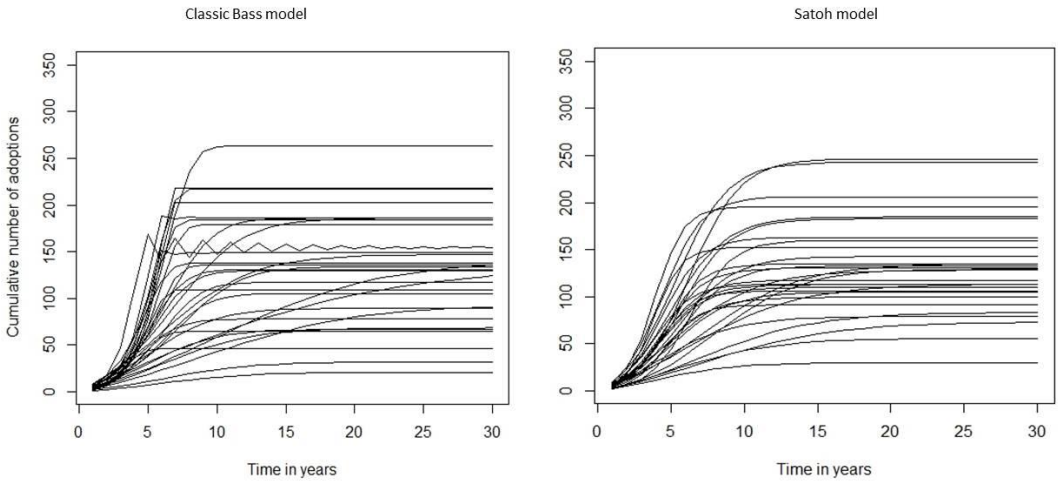


Figure 6.11 Five likely shapes of diffusion curves using the Satoh approach



Randomly sampling 30 curves from both the Satoh and the classic Bass model shows that sampling from the classic model can result in unrealistic diffusion curves (Figure 6.12). There is a few curves in the left hand side of Figure 6.12 that appear to have a peak with a following decline. One curve is obviously oscillating.

Figure 6.12 Thirty sampled diffusion curves from classic and Satoh model



6.3.1.2 Cumulative diffusion of EIS with and without research

The results from the elicitation suggest that there would be an increase in uptake with further research. For example, the attainable number of adoptions increases with further

research, although uncertainty is larger than for the estimate without research (Figure 6.13). Figure 6.13 shows density plots of the elicited quantities that are used to generate the Bass model. The left hand side of Figure 6.13 shows the estimated quantities of interest for the counterfactual scenario in which no further research is conducted. The right hand side shows the probability distributions of the same quantities conditional on both identified relevant trials being conducted. Table 6.3 shows the means and standard deviations of all the elicited quantities.

The mean estimate of the attainable number of adoptions is relatively small. Given that there are 231 obstetric units in England and that some units would be expected to purchase more than one device (large units but also units that want to have a back-up when one of the devices is in maintenance), the number of 140 adoptions is not large. The first year will only see a very small number of adoptions, possibly by those units who developed the technology or have research grants for it. This was the rationale given by the experts for the number of adoptions being very small at a mean of five adoptions or ten adoptions if more research was conducted. More explanation will be given in the section on the individual elicitations. More research will make itself known more quickly. This was the underlying reasoning for the number of years at which the maximum number of per period adoptions is reached: it was estimated to be at a mean of six years, or five years with more research.

Research on the relative treatment effect of progesterone compared with no treatment would contribute more to the overall increase in adoptions caused by further research than another study on the predictive ability of EIS (shown in Table 6.3 and Figure 6.14). This is expressed through the quantity κ being estimated at a mean of 61%, although there was considerable uncertainty surrounding it. Even when EIS is purchased, it will only ever be used on an expected 60% of patients (expressed by the quantity π). Experts commented on there being logistical difficulties that are hard to overcome, doctors' scepticism and the cost of disposables that would affect usage after purchase.

Figure 6.13 Elicited probability distributions of diffusion quantities for each expert and linear pool

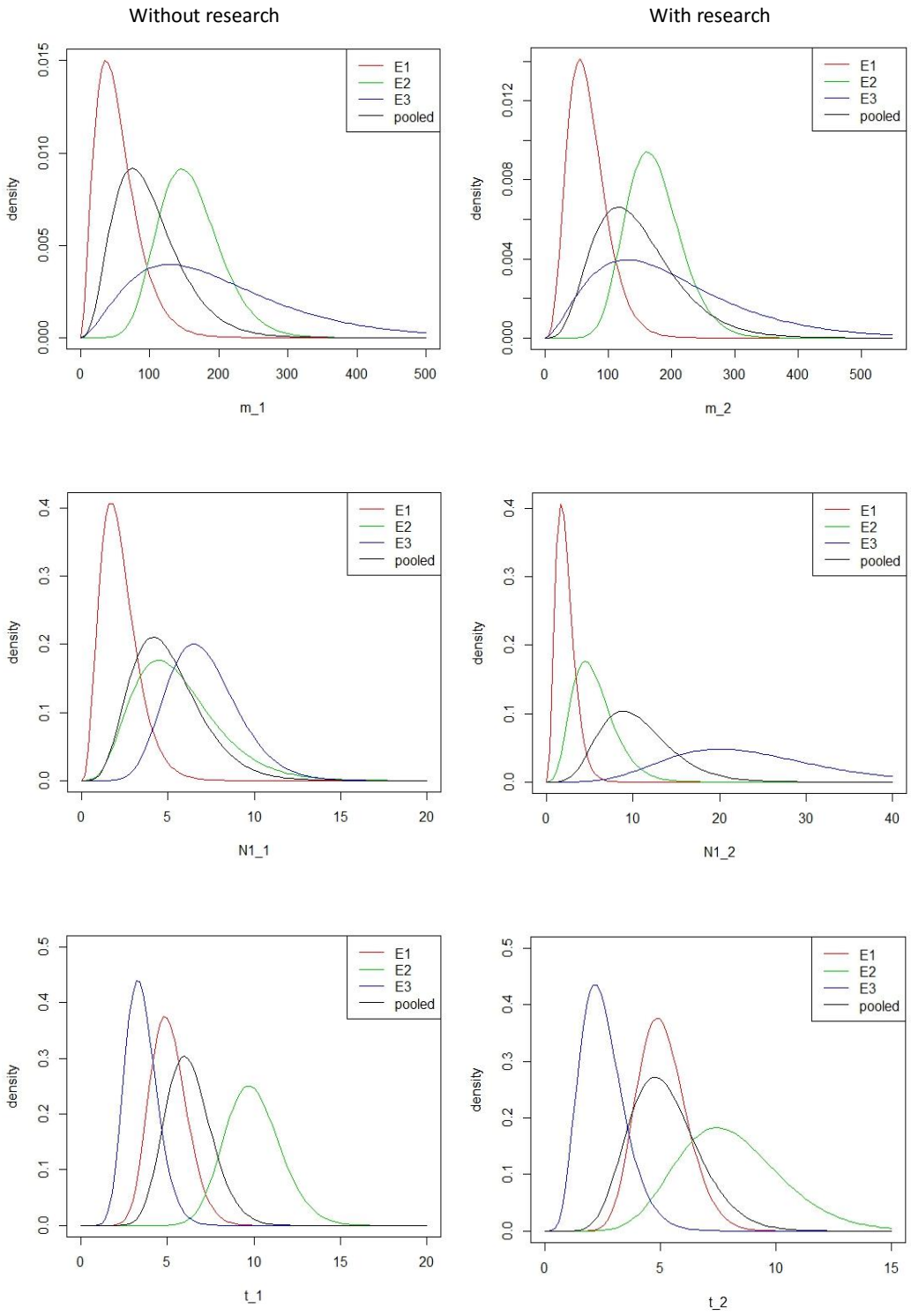


Figure 6.14 Elicited probability distributions of kappa and pi for each expert and linear pool

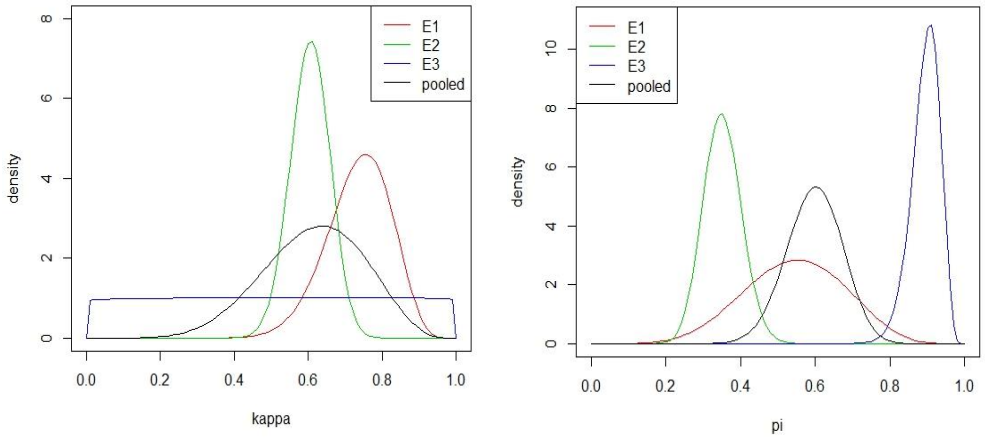


Table 6.3 Pooled elicited quantities (mean, standard deviation (SD), median and 95% credible intervals (CIs))

Without further research					**With further research**				
Variable	Mean	SD	Median	CI	Variable	Mean	SD	Median	CI
m1	140	69	129	(39-304)	m2	146	66	136	(46-301)
n1	4.99	2.04	4.71	(1.8-9.7)	n2	10.50	4.12	9.96	(4.0-20)
t1	6.26	1.34	6.16	(3.9-9.1)	t2	5.21	1.52	5.06	(2.7-8.6)
The contribution of progesterone trial (κ) to increase in diffusion and utilisation (π)									
Variable	Mean	SD	Median	CI	Variable	Mean	SD	Median	CI
κ	0.61	0.14	0.62	(0.33-0.85)	π	0.60	0.07	0.60	(0.44-0.74)

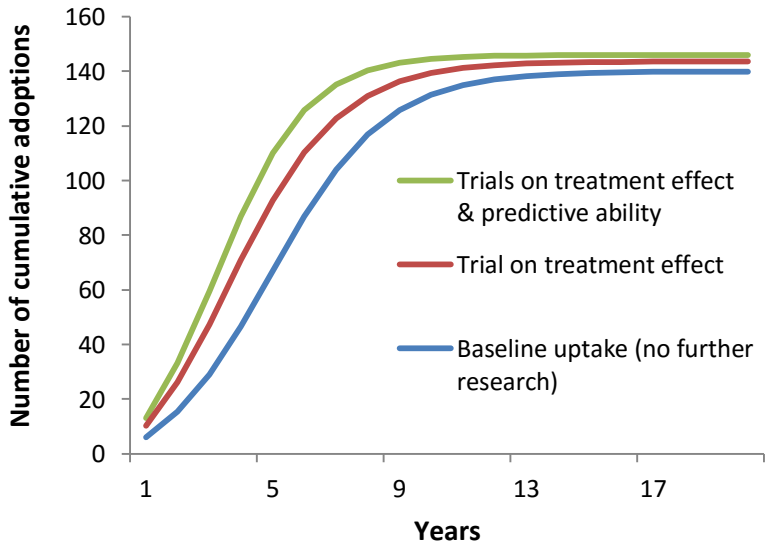
The resulting means for Bass model parameters *p*, *q* and *m* with and without research obtained from the simulation that used the Satoh method are presented in Table 6.4.

Table 6.4 Values of simulated Bass model parameters (mean, SD, median and CIs)

Bass model parameter	**Without further research**				**With further research**			
	Mean	SD	Median	CI	Mean	SD	Median	CI
p	0.018	0.020	0.026	(0.005-0.083)	0.040	0.038	0.049	(0.013-0.157)
q	0.491	0.176	0.482	(0.2-0.806)	0.537	0.136	0.524	(0.313-0.826)

As per the definition of the Bass model, the cumulative diffusion of EIS exhibits an s-shaped growth as can be observed from Figure 6.15. This shows the pooled mean cumulative diffusion curves with and without research. The blue and green lines were elicited and the red line was obtained by applying κ to the difference of with and without research adoptions in each period, as was discussed above.

Figure 6.15 Cumulative mean number of adoptions without and with the availability of further research evidence



Comparing the graph in Figure 6.15 with the means of the elicited quantities in Table 6.3 suggests that the method of approximating parameters *p* and *q* with the elicited quantities works relatively well although there are deviations from the elicited mean quantities. The approximated curve shows a number of first year adoptions of $N_1^* = 6.08$ units which increases to $N_1^* = 13.01$ units with further research (compared to $N_1 = 5$ and $N_1 = 10.5$ in

the elicitation). Note that the asterisk represents that these are approximated values, while the elicited values are represented by plain variable names, for instance, N_1. The point of inflection occurs at $t'^* = 5$ years without further research and can be brought forward to $t'^* = 4$ years with further research ($t' = 6$ and $t' = 5$ years in the elicitation).

Figure 6.15 shows that it takes approximately 14 years for EIS to be adopted by all the potential attainable adopters; with further research this is shortened to 10 years. With growth of the curve slowing down in the last periods, this means that 90% of the adoptions are achieved after nine or six years, without or with further research, respectively. Half of the adoptions are achieved at six or four years, respectively.

The classic Bass model would have shown a slightly better fit in the mean values, but it is in the extreme values that *p* and *q* could adopt, where this method fails. For example, the approximated curve would have led to ${N_1}^* = 6.18$ without further research and ${N_1}^* = 10.28$ with further research, a more accurate fit with the elicited values in Table 6.3. The point of inflection would have been at $t'^* = 6$ and $t'^* = 5$ years, almost a perfect fit with the elicited values and certainly a better fit than provided by the Satoh approach. Despite this better fit near the mean, the classic Bass model could not be used going forward because unrealistic diffusion scenarios could not be prevented, as was shown in Figure 6.2.

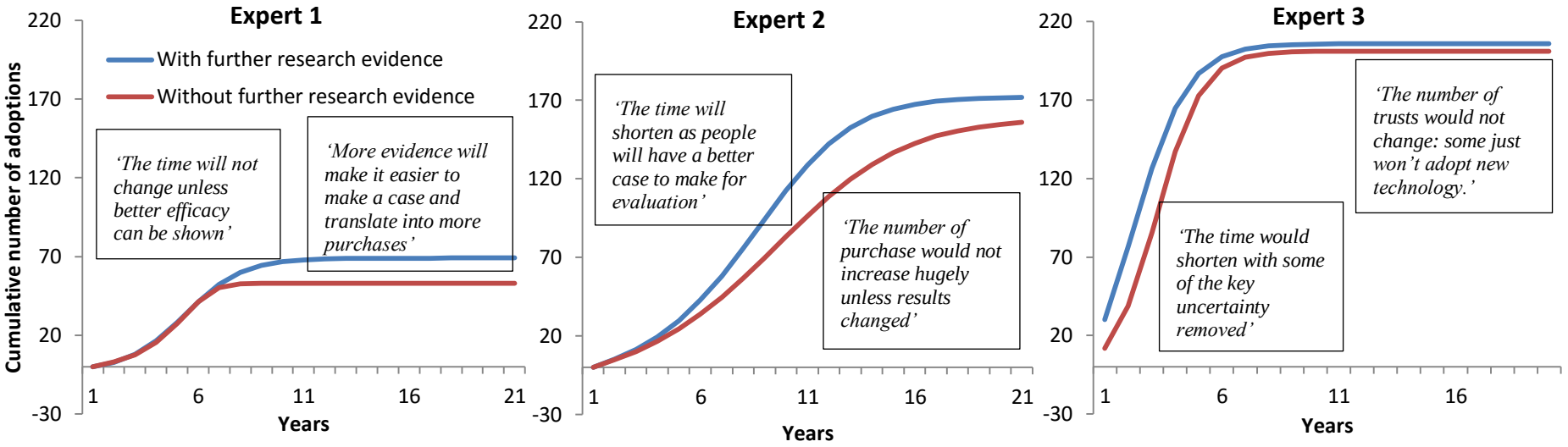
As highlighted above, the number of adoptions does not increase much with additional research. In fact, the cumulative number of adoptions only changes by six adoptions, or by four per cent. Diffusion appears to speed up with more research, with a reduction in the time to the maximum number of adoptions of one year or four percent. The number of adoptions in the first year is what changes the most, approximately doubling from six adoptions to 13 with further research. The relatively small impact of research on diffusion in this case study is surprising given that the lack of research was identified as one of the barriers to diffusion in EIS. Possible reasons for this result will be discussed in the section on the individual elicitations where I present qualitative insights provided by the experts, as well as in the discussion.

6.3.2 Individual diffusion curves and rationales

Heterogeneity was observed in the individual diffusion curves elicited from each expert. These variations manifested themselves in the experts' beliefs about the total attainable number of adoptions and the time that it would take until they are achieved as well as in the effect of further research on diffusion. These differences are going to be discussed below.

The diffusion curves resulting from the elicited mean parameter values with and without conducting further research are shown in Figure 6.16. Experts' rationales for changes in the total number of purchases and the time at which purchases start to decline have been added to the figures but are explored more in depth in the following.

Figure 6.16 Individually elicited mean diffusion curves from 3 experts for with and without research scenarios



6.3.2.1 The attainable number of adoptions (m)

Experts' estimates of the attainable number of adoptions varied significantly (Table 6.5). Expert 1 predicted the lowest number of attainable adoptions both with and without further research evidence. The rationale for this low estimate was that, in their opinion, the positive predictive value of EIS was relatively small (currently estimated at 63%). The very good negative predictive value (currently estimated to be at 99%) could not offset this and this expert claimed that apart from reducing uncertainty, research would have to show a significantly improved predictive ability of EIS in order for the number of purchases to increase. In their opinion, the main driver for the adoptions that could be achieved despite this was the fact that EIS would alleviate the dependence on ultrasound.

Table 6.5 Mean (95% credible interval) estimates of attainable number of adoptions

Expert	**1**	**2**	**3**
Without research	54.2 (10-150)	158.8 (30-230)	204.4 (30-410)
With research	68.7 (20-150)	174.2 (60-250)	205.5 (30-410)

Expert 2 on the other hand thought that the positive predictive value was good and that there was a lot of potential in the device. Reducing uncertainty through research would increase the number of attainable adoptions but, similarly to expert 1, expert 2 thought that improved results would have an even larger effect on it. Expert 2 furthermore thought that there was always going to be trusts that will not adopt the new technology and that further research could not change this. This innovation-adverse behaviour exhibited by certain trusts could only be changed through national guidance, in their opinion.

The third expert estimated the largest values for the attainable number of adoptions despite expert 3 thinking that the positive predictive value was not that good. However, expert 3 stated that the strong negative predictive value would be of great help providing certainty for women testing negative. EIS was also perceived to be less costly and easier to apply than ultrasound. Expert 3 stated that the number of attainable purchases would not change with the availability of more research, as there are some trusts that do not adopt new technologies for reasons of budgets and lack of knowledge.

6.3.2.2 The time at which the number of adoptions starts to decline (t')

The time it takes to achieve the maximum per period number of adoptions was estimated to be quite different depending on which expert was asked (Table 6.6).

Table 6.6 Mean (95% credible interval) estimate of number of years up to the peak of the per period diffusion curve

Expert	**1**	**2**	**3**
Without research	5.13 (3-8)	9.88 (7-13)	3.53 (2-6)
With research	5.11 (3-8)	8.16 (5-11)	2.57 (1-5)

Expert 2 was the one with the largest estimate. Their rationale for it taking so long was that the NHS is known to move slowly. The time would be shortened by more than a year with more research available as people would be convinced more easily and it would be easier to make a business case. The first expert found a shorter time to the peak of adoptions. Expert 1 explained this with the relatively low success they thought EIS would have. The expert's rationale was that those who adopt would do it relatively fast, although they did concede that it takes time until a new device gets talked about. Expert 1 also believed that further research showing the same results would not change the speed of diffusion, which could only be improved through evidence that shows improved predictive ability or treatment effect of progesterone.

Expert 3 provided the shortest estimates. This coincided with them being the expert with the highest attainable number of purchases suggesting that they believed in a much greater success than the other participants. Their rationale was that the problem of preterm deliveries is viewed as a great clinical need, which people are desperate to address. They thought that further research would help shorten the time further by reducing some of the uncertainty.

6.3.2.3 The number of adoptions in the first year (N1)

Two experts thought that no change would occur in the number of adoptions in the first year when further research was conducted (see Table 6.7). Both their rationale was that those early adopters would be people involved with the research and that this would not change

with the availability of further research. Expert 1 thought the initial purchases to be very low due to the purchasing process taking very long, a year at minimum. Expert 2 thought the same but added that early meetings and presentation of findings at conferences could lead to higher numbers in the first year.

The other expert, however, believed that there was going to be a major change (expert 3). Expert 3 believed in the largest number of purchases in the first year and further believed in a stark increase with the availability of further research evidence. The rationale provided was that a greater number of research studies would enter public consciousness sooner. Expert 3 also thought that initial purchases would largely stem from those units involved with research or associated with them. Expert 3 shared the other experts' views that the purchasing process can be very lengthy and also voiced the opinion that people are generally reluctant to change and that the numbers depended on costs and presentations at conferences.

Table 6.7 Mean (95% credible interval) estimate of number of adoptions in the first year

Expert	**1**	**2**	**3**
Without research	2.25 (0-5)	5.7 (2-15)	7.1 (2-10)
With research	2.2 (0-5)	5.5 (2-15)	23.7 (10-40)

One avenue worth considering for synthesising these different estimates is an approach by which the anonymised experts' estimates and underlying rationales are fed back to all the experts and by giving experts the opportunity to adjust their views. The advantage of this could be that experts may reach a more informed estimate by considering other experts' views. The drawbacks are the additional time and potentially resource requirements on the part of the researcher and the experts. Furthermore, I am not aware of any problems of over-confidence resulting from this approach, but the possibility would need to be considered.

6.4 Reflection on the exercise

As Soares et al. (2011) stated, it is impossible to assess the face validity of an elicitation exercise as the expert's true beliefs are unknown. What can be done is reflect on the exercise as well as on feedback provided by the experts. The present elicitation exercise had certain characteristics that perhaps made it especially challenging. The topic of the exercise, future diffusion of a not yet existing health technology, may have lain out of the comfort zone of the participating experts. While experts were selected based on the belief that they were suitable candidates for providing estimates of diffusion, diffusion was certainly not something they could be expected to think about in their daily lives. The fact that future diffusion is dependent on many other variables further increased the complexity of the exercise.

These difficulties made it necessary to conduct relatively lengthy briefing. This briefing included, as mentioned above, agreement with regards to assumptions made on the variables influencing diffusion. The values for the predictive ability of EIS and the treatment effect of progesterone after EIS screening, subject to uncertainty, were only two of many assumptions made. It was impossible to share assumptions on all the factors that may influence future diffusion (as identified in the previous chapter) in the briefing. This may have led to experts making different assumptions on some of the factors, such as costs, and some of those may have been made subconsciously.

It is therefore especially hard to evaluate the effectiveness of such an elicitation. Experts' calibration may be related to their knowledge but also to assumptions they made regarding the future environment which may result in experts' estimates being more accurate in certain settings that may or may not realise in the future. While calibration cannot be checked at this point in time, there was suggestive evidence for experts using different assumptions to obtain their estimates of diffusion. This became evident when examining the rationales for their beliefs in certain parameter values. However, this uncertainty was presumably captured in the probability distributions.

In one case, the same rationale translated into quite different probability distributions for a certain parameter. For the number of purchases in the first year, expert 3's estimate significantly diverged from the other experts despite them stating more or less the same rationale. This may be an indication that it was difficult for the experts to think about this particular quantity, or that experts made different underlying assumptions that they did not make explicit. This is not surprising given that these prior distributions characterise the

experts' beliefs in the face of limited knowledge, which means that their distributions can only converge as more evidence becomes available.

Interestingly, different rationales could also lead to similar quantitative beliefs. The time until the number of purchases declined was estimated to be shorter by two experts who had opposed views on the overall success of EIS. Expert 1 thought that it would not take long for EIS to reach its peak, given the low number of attainable purchases. In contrast, expert 3 believed in a very fast product growth given the overall success of EIS. This suggests that the time to the peak number of adoptions does not indicate a successful or unsuccessful technology. Another insight is that this quantity was unlikely to be strongly correlated with the attainable number of purchases.

Although the process encouraged the participants to voice all their rationales and I made sure to record them, it cannot be guaranteed that all of these rationales and underlying assumptions were captured. This may have to do with the fact that experts themselves might not have been aware of the assumptions they were making and that they were therefore not making them explicit.

The above observations are supported by one expert's feedback on this exercise:

> *'It was difficult to predict future events like this but I'm relatively confident with the estimates given that the conditions set for the task remain true.'* (Expert 1)

The others merely stated that they found the task interesting and in one case, the software clever. All of them were happy to participate and understood the task well after the training was conducted. The experts also commented on difficulties they had with the following:

- The use of tertiles: all experts initially thought of tertiles being closer to the upper and lower limits of the range than to the median; suggesting a bimodal distribution would make the best fit. Showing them the fit of their distribution in a histogram and how the choice of tertiles affected the shape of the distribution was one step to clarifying their meaning. It was also explained that the closer the tertiles were to the median, the more likely the values close to the median became and the narrower the resulting distribution. It was explained that tertiles should be chosen such that there was a 33% certainty that the value fell into that range between the two tertiles. Even after that explanation, the participants found it difficult to come up with numbers for the tertiles and relied more on the visual representation of their resulting distribution. An iterative approach, by which participants shifted tertiles and

checked the effect on the visual, gave them the opportunity to achieve the shape of distribution they wanted, that is the desired uncertainty and skew. In some cases, experts realised that their range might have been too narrow. In such cases, the range was widened. The use of quartiles may be more intuitive and should be considered in future studies.

- Difficulty with underlying assumptions: experts found it difficult to make judgements considering the lack of information there was on all the influencing factors to uptake in the future. Experts were encouraged to widen their ranges if they felt great uncertainty. One particular assumption caused difficulty for one expert: the one that the current EIS trial would remain the only one. This was seen as unrealistic by this expert but I encouraged them to keep that assumption for the purpose of this exercise. The lack of information on the quality of the current EIS trial caused a problem for the elicitation of the relative importance of the progesterone trial over another EIS trial. One expert felt unable to give a judgement because of that. This needed addressing in the data analysis: a flat beta(1,1) distribution representing this expert's opinion was thus used in the linear pool.
- Thinking about the number of years at which the number of purchases peak: this was viewed as the most difficult quantity to estimate. A conversation on potential rationales for different lengths of time helped clarify thoughts. Also, the possibility of reflecting their uncertainty through selecting a wider range or through the choice of tertiles was highlighted again. Peculiarly, experts did not choose particularly weakly informative distributions. It is hard to tell whether this might have been due to over-confidence (it is important to bear in mind that expert knowledge is fallible (O'Hagan and Oakley, 2004)) or whether this represented their true beliefs. Over-confidence would have, at least partially, been addressed by linear pooling of the three experts' beliefs (Johnson et al., 2010).

The above shows that it is difficult to develop a way of eliciting diffusion that minimises bias and avoids the use of certain heuristics and assumptions but that appropriate preparation, briefing and training have the potential to support such an exercise.

6.5 Exploring the impact of trial results on diffusion

In this chapter, I examined the impact of reduction of uncertainty through the generation of research evidence on diffusion. For this task it was implicitly assumed that study results

of both simulated studies would be similar to existing evidence (experts were asked to bear some uncertainty in mind). Given that experts diverged significantly in their estimates of attainable number of adoptions (m) and the time to the peak number of adoptions (t'), and given that the experts expressed differing views on the quality of the existing evidence, it can be hypothesised that experts' estimated diffusion curves would take different shapes conditional on the data generated by the respective studies. For instance, if EIS was shown to have a better predictive ability than that so far observed (in terms of positive and negative predictive values, for example), diffusion may be quicker or a greater number of attainable adoptions could be achieved. If, in another example, the relative treatment effect of progesterone versus no treatment was much worse after EIS scans than after CL scans, then diffusion may turn out to achieve fewer attainable adoptions.

This means that diffusion estimates should not only be elicited for the counterfactual with and without research scenarios, but also for different possible study results. Continuing with the example on the progesterone trial, one could elicit three different scenarios: that progesterone treatment effect after EIS screening turns out to be similar to that predicted for CL scans (for instance, within the two middle quartiles of the prior distribution), that it turns out to be significantly worse (in the lowest quartile of the prior distribution), and that it turns out significantly better (in the uppermost quartile). Of course, a more detailed breakdown of different research study outcomes is desirable (for example, quintiles), but elicitation exercises only work well with a small number of elicited quantities. Assuming that the value of one well-defined trial is going to be assessed, elicitation of a counterfactual without research diffusion curve is required, as well as elicitation of three with research diffusion curves, resulting in 12 quantities to be elicited. Using quintiles would drive this up to 18 quantities, which may lead to exceeding the time that experts' are able and willing to allocate to an elicitation meeting.

When diffusion is elicited conditional on study results, it may also be worthwhile to provide information on the potential cost-effectiveness associated with each of the possible sets of study results, because knowledge of cost-effectiveness may also influence the diffusion curves.

6.6 Discussion

6.6.1 Findings

In this chapter, I presented a method for eliciting expert beliefs about technology diffusion using adaptations of the Bass model of diffusion. This method was used in estimating diffusion of EIS for use in PTB screening. I demonstrated an approximation method for converting experts' beliefs on observable quantities into Bass model parameters using two different specifications of the Bass model: the Satoh adaptation and the classic Bass model. The Satoh specification of the Bass model performed better than the classic model with respect to the resulting parameter space, enabling the generation of s-shaped diffusion curves. The classic model reflected experts' beliefs slightly better near the mean estimates but caused problems in extreme parameter values that could not be overcome. The model parameters *p* and *q* were shown to be negatively correlated and also correlated with parameter *m*. The correlation is a feature of inverting a non-linear function.

Results for EIS showed that the attainable number of adoptions was relatively low. While no authoritative figure on the desirable number of adoptions exist, an average of two devices per obstetric unit in England could be plausible, resulting in a tentative value of 462 desirable adoptions. A mean of 140 attainable adoptions without further research evidence is therefore low, resulting in 30% of desirable adoptions only. The experts' main rationales were: the predictive ability of EIS being less good than the experts would like it to be at present and the general adversity to change that was thought to be present in the English NHS. It would take approximately 14 years for EIS to reach all attainable adoptions without, or 10 years with further research evidence available. Once EIS is purchased by the hospital, only an expected 61% of patients would be offered the screening with it. Assuming that all hospitals saw the same number of patients, this would result in only 5,000 of 27,000 patients being offered EIS screening. As stated previously, I will not use utilisation for the remainder of the thesis. However, the decision-maker and the developer of EIS may wish to bear this low utilisation in mind when devising implementation strategies. These findings do not take into account the potential differences in diffusion curves conditional on different trial results.

6.6.2 Relevance

This study adds to current knowledge by offering an approach to obtaining probabilistic, technology-specific predictive estimates of diffusion even before technology launch. The

reviews in Chapter 2 showed that predictive estimates of diffusion are rarely used in health economic evaluation. In Chapter 3, the reviews identified no existing methods for obtaining probabilistic health technology diffusion estimates prior to technology introduction but did highlight that an expert judgement exercise should use formal diffusion models, such as the Bass model, as the underlying structure. The method commonly used for predicting future diffusion included guessing by analogy, which was not deemed suitable in health technologies for a lack of data on analogous technologies, and because it would ignore the technology-specific factors influencing uptake. The method of elicitation of experts' beliefs together with specifications of the Bass model used in this study therefore fills this research gap. It is useful in health technologies, because there usually is not sufficient data for a guessing by analogy exercise, but it can also be useful for other types of technologies because this method allows taking into account the very distinctly individual environmental and technology-specific conditions in estimating diffusion. Although the Bass model of new product growth itself is not new, its use has been limited to predict future uptake when data on the first few periods were available or to guessing by analogy studies.

The contribution of probabilistic estimates of diffusion for EIS as a PTB screening tool may prove important for further cost-effectiveness and value of information and implementation analyses, examples of which will be presented in Chapters 7 and 8. The presented approach enables researchers to obtain better representations of uptake patterns and thus may help direct public funds to the most efficient implementation strategies. This approach could also be of interest to manufacturers who may want to assess the value of investing into technology development at an early stage, as technology uptake is strongly linked to potential profit through revenues.

6.6.3 Strengths and limitations

The strengths of this approach relate to the representation of experts' uncertainty. Experts could also take factors influencing future uptake into account, leading to the best guess that we can obtain at present. Experts were encouraged to provide insights into their thought process and describe those factors that influenced their estimates the most. These additional data provided the opportunity of analysing the implicit assumptions experts were making.

The use of the Bass model enabled me to predict future diffusion with a minimum number of elicited quantities. Rather than eliciting uptake levels for every year, only three points sufficed to obtain uptake dynamics that are coherent with the theory of diffusion of

innovations and that have been shown to more or less accurately predict diffusion of other technologies. Reducing the number of elicited quantities is important: not only is recruitment harder with increased meeting lengths, but experts' calibration may also suffer as the task goes on.

Limitations of this approach were mainly related to the complexity of diffusion and its dependence on a diverse set of factors. In particular, the present elicitation exercise did not consider the different possible study results. However, during this exercise, experts flagged up that diffusion would be different conditional on different study results. This means that a major part of the uncertainty associated with diffusion is potentially not reflected in the elicited distributions. I have therefore proposed a way of structuring the elicitation exercise to account for different diffusion curves conditional on study results in Section 6.5.

A caveat of the use of the Bass model is that the joint parameter space may contain implausible values because I was unable to multivariately elicit beliefs about the observable quantities. Ideally, these correlations should have been quantified using multivariate elicitation but this was beyond the scope of this research because of the increased difficulty and meeting length of a multivariate elicitation. Another drawback of using the Satoh approach to inverting observable quantities into Bass model parameters was that the resulting range for parameter q was smaller compared with the classic Bass model parameter. Unfortunately, it was not possible for me to assess whether this led to introduction of bias, let alone quantify the bias.

Difficulties were noted in the elicitation of certain quantities: while the quantity m and N_1 (that is, the number of attainable adoptions and the number of adoptions in the first period, respectively) were intuitively easy to grasp, respondents found it more difficult to come up with an estimate of t' (the time until the per-period adoptions start to decline). The quantity κ (the proportion of uptake increase caused by the study on the treatment effect of progesterone) posed another problem that might have been caused by the parameter's definition, which was confusing to respondents. The assumption that the effects of different types of research on diffusion were additive was another limitation and will limit the use of κ in further analysis. Ideally, the contribution of each study would have been elicited separately.

It should also be highlighted that elicitation of future diffusion, not unlike elicitation in general, requires a significant amount of preparation and planning. The particular challenges in this elicitation lay in briefing experts on uptake of health technologies and explaining the elicited quantities as well as presenting the different research findings relevant to this

technology. As well as gathering the information for the briefing, time should be taken to phrase questions asked in the elicitation and practising the use of software as well as providing training to experts. Conducting elicitation to inform an uptake model may thus not be feasible in all research settings.

One of the greatest difficulties was associated with finding a balance between guiding the experts in the assumptions they were making about the underlying conditions of diffusion and giving them absolute freedom. The former could produce probability distributions that are too narrowly defined; the latter could lead to weakly informative probability distributions including unrealistic scenarios. I decided to set out conditions, but even these few conditions meant lengthy briefing of experts before the elicitation was started. Another drawback of setting out these conditions was that the resulting diffusion curves may contain a bias or false certainty. I do not believe that briefing was avoidable for the simple reason that experts were not familiar with EIS at all. I also found that experts may refuse to provide any estimates if they feel too uncertain about pivotal factors that would influence their beliefs. This is an area for further research in the field of elicitation.

With the trial still ongoing, no results were available to present to the experts and it had was assumed that estimates of the predictive ability obtained from the pilot would hold true. This may compromise the usability of the resulting diffusion curves for health economic decision-making because the evidence that will be produced in the ongoing trial may alter the experts' views of future diffusion. It can be argued, however, that experts' estimates of diffusion reflected uncertainty, also about the trial outcomes. In Section 6.5, I outlined a more sophisticated approach to reflect the impact of different trial results on diffusion. The fact that the elicitation was conducted early on also showed that it is possible to elicit diffusion estimates at an early stage of product development, which may be relevant in instances in which the manufacturer wishes to find out about potential diffusion of a technology early in product development to help decide whether to pursue that technology or not.

The use of the tertile method proved difficult for the experts. While elicitation of the range and the median did not pose any problems, experts found it hard to provide opinions on tertiles, possibly due to their unfamiliarity with probability distributions. I addressed their questions by showing visuals of histograms or density curves, showing examples of different tertiles. The main benefit of using tertiles was that no intervals had to be pre-specified and the expert therefore had greater flexibility in deciding on the shape of their probability distribution with the estimation of very few points. This greater flexibility came at the expense of simplicity associated with fixed interval estimation methods. Since this initial

lack of understanding could be overcome, it does not present a problem for the task in general, but it is worth bearing this trade-off of flexibility versus simplicity in mind when designing future studies. The use of quartiles may be more intuitive for participants and should be explored in further research.

It would have been desirable to show the effect of elicited parameters on the diffusion curve during the meeting as a means of providing feedback and to improve the fit of probability distributions with experts' opinions. It was not possible to do this in this study as running solver with the elicited quantities to yield the Bass model parameters and produce a mean diffusion curve would have resulted in exceeding the available time for the meeting.

6.6.4 Further research

Scope for further research lies in addressing the limitations of this study, particularly the elicitation of different diffusion scenarios conditional on trial results as was outlined in Section 6.5. Furthermore, the correlation of parameters could be explored through multivariate elicitation, poses its own set of challenges that was beyond the scope of this study. Whether experts are truly unable to provide the covariance structure should ideally be explored through performing research on this but this may not be feasible because of the increased lengths of elicitation meetings.

It would be interesting to repeat this study once the first trial results are available and examine any changes to diffusion with potential changes to the predictive ability of EIS. I did not examine the potential increase in uptake that could be caused by better effectiveness results for EIS because this is not directive for action as, presumably, the device's predictive ability cannot be improved. This study revealed, however, that the device's predictive ability, especially its positive predictive value, may pose a significant barrier to implementation. Eliciting uptake estimates conditional on what if scenarios with different levels of efficacy could thus be useful to manufacturers who wish to make a decision on whether to develop a technology further and for obtaining a more accurate and complete picture of the value of implementation that appropriately reflects uncertainty around implementation. In this case study, predictive ability may indeed be improved, as EIS has been developed further since the pilot study, on which the positive and negative predictive values used in this exercise were based.

6.6.5 Conclusion

In conclusion, I recommend elicitation of expert opinions to inform the Satoh modification of the Bass model of diffusion as a way of obtaining probabilistic, dynamic and technology-specific estimates of uptake. The estimates of diffusion obtained through application of this method to EIS for use in PTB screening can be used for cost-effectiveness and value of information and implementation analyses, although further exploration of different study results and their impact on diffusion is desirable.

CHAPTER 7. THE DYNAMIC EVII FRAMEWORK APPLIED IN AN ECONOMIC EVALUATION OF EIS FOR USE IN PTB SCREENING

7.1 Introduction

The aim of this chapter was to illustrate the dynamic expected value of information and implementation framework that was developed in Chapter 4 in the exemplary case study of electrical impedance spectroscopy (EIS) for preterm birth (PTB) screening. Background on the case study technology, the therapeutic area and health need was provided in Chapter 1. The latest trial evidence on the predictive ability for EIS was not available upon writing this thesis because it was still ongoing, and any results are therefore preliminary and based on pilot study data. I drew on insights from the qualitative study on potential implementation strategies presented in Chapter 5. I also used the estimates of counterfactual with and without research implementation dynamics obtained from the elicitation study described in Chapter 6. The flow of the thesis and the place of this chapter in it are shown in Figure 1.2 in Chapter 1.

In this chapter, I wish to address the following objectives:

1. To demonstrate the use and usefulness of the concepts related to the dynamic expected value of information and implementation in the context of HTA.

2. To compare static implementation estimates with the elicited diffusion estimates with respect to their impact on health economic outcomes and implications for decision-making.

This chapter will follow these steps: I describe the PTB screening model that I developed for the economic evaluation of EIS in Section 7.2. The dynamic EVII framework that was developed in Chapter 4 is applied in the PTB screening model in Section 7.3. I present results of the decision-analytic model for EIS and compare static with dynamic results in Section 7.4. I conclude with a discussion in Section 7.5.

7.2 The preterm birth screening model

This section describes the conceptualisation of and data inputs used for the economic model of EIS for use in PTB screening. For the model conceptualisation, I followed the steps outlined by Gray et al. (2011) to build a decision-analytical model and guidance by Squires (2014) to conceptualise the model.

7.2.1 Defining the research question

7.2.1.1 Review of existing health economic models for PTB screening

I started model development with a review of other economic evaluations for PTB screening as Squires (2014) highlighted the potential helpfulness in reviewing previous economic evaluations in the therapeutic area. With EIS still being under development at the time of writing, there was no economic evaluation of EIS for use in PTB screening. There were, however, two existing economic evaluations of screening interventions combined with treatments in the therapeutic area of preventing PTB: one HTA by Honest et al. (2009), and one cost-effectiveness analysis by Werner et al. (2011) that was updated by the authors in a later publication (Werner et al., 2015). I briefly discuss these two economic evaluations in terms of their strengths and limitations for use in assessing EIS:

The HTA report by Honest et al. (2009) assessed a large number of different screening tests (22) and interventions (40) for their effectiveness in predicting and preventing preterm birth in asymptomatic women and women with symptoms of preterm labour. Cervical length ultrasound scans were among the screening tests and progesterone was assessed as an intervention to prevent preterm birth. The model structure used was a decision tree model for both asymptomatic and symptomatic women. Four options (test no one and treat all, test all and treat no one, test all and treat only with positive test and test all and treat all) were compared to test no one and treat no one. The model outcome in asymptomatic women was the cost per threatened PTB avoided; in symptomatic women it was the cost of preterm birth avoided. Effectiveness was thus not measured in QALYs but only in threatened preterm birth or actual preterm birth avoided. The average cost associated with a woman tested and treated in the symptomatic analysis was the cost that screening and treatment of asymptomatic women is attempting to avoid – hence the symptomatic analysis was necessary for the analysis in asymptomatic women.

The economic evaluation performed by Honest et al. (2009) was useful for the purpose of this study in that it provided a systematic review of the effectiveness and costs associated with different screening and treatment interventions. The fact that it adopted a very short-term time horizon by ignoring costs and QALYs associated with the future life-times of the affected babies is a limitation. The lack of QALYs also limits the EVI analysis as there is no accepted monetary valuation for the outcome measures used in their analysis.

Werner et al. (2011) assessed the cost-effectiveness of a single routine transvaginal cervical length measurement at 18-24 weeks of gestation followed by daily vaginal progesterone treatment for those women identified at risk against no screening and no treatment. Like the study by Honest et al. (2009), the model has a decision tree structure. This analysis used costs and QALYs as the model outcomes. There were only three possible health states that were associated with four different lengths of gestation: full health, severe disability and death. The advantage of this analysis was that it took a life-time horizon and quality of life into account. However, evidence used on the effectiveness of progesterone and the predictive ability of cervical length (CL) scans was based on single trials rather than a systematic review and evidence synthesis and health outcomes were crudely modelled reflecting only three health states.

The study by Werner et al. (2011) found cervical length scans combined with progesterone treatment to be a dominating strategy in comparison to test nobody and treat nobody. In a subsequent revision of some of their input parameters (Werner et al., 2015), the authors still found this strategy to be cost-effective, though not dominating. The same strategy did not come out to be the most cost-effective in the model by Honest et al. (2009). This difference in model results appears to be caused by the difference in the time horizon that is considered for the studies as well as by the costs associated with preterm birth. Both studies used a decision tree model, which is appropriate to answer the research question (see justification below in step 2). Werner et al. (2011) considered the impact of four different lengths of gestation on costs and health outcomes while Honest et al. (2009) only considered two different lengths of gestation and their impact on costs.

In summary, the two studies differed in terms of the use of evidence, modelled health states, and length of time for which costs and health outcomes were modelled. Both studies could be helpful in providing data for the present economic model, although some of the input data needed updating.

7.2.1.2 Research questions for EIS for use in PTB screening model

The present model aimed to estimate the costs and effects of EIS screening followed by progesterone treatment in women with a singleton pregnancy and a history of PTB or mid-trimester loss; and to compare these outcomes to those of comparator screening interventions followed by progesterone treatment. PTB was defined as birth before 34 weeks of gestation, at which time the threat to the baby is the most significant (Werner et al., 2011). The comparator screening interventions used in this model were cervical length ultrasound scans and no screening. The comparators CL scans and No screening had been identified as the most relevant comparators in the interviews with experts and based on communication with the developers of EIS. The interviews also identified progesterone as the only likely prevention strategy for those who tested positive. A no screening, treat all strategy was deemed improbable to occur in practice and was therefore not assessed. In contrast to the study by Honest et al. (2009), all strategies were only assessed in high risk women, following the identification of parous women at risk of preterm birth based on previous history of at least one preterm birth.

The analysis was conducted from the payer's perspective, and within the jurisdiction of England, therefore taking the viewpoint of the England NHS. If also a personal social services (PSS) perspective was adopted, EIS would likely be more cost-effective because fewer personal and social services would be needed in the case of averted PTB. It is also worth noting that decision-makers at a local level may be more interested in the cost-saving in the pregnancy and delivery pathway than the lifetime costs. A scenario analysis from local level decision-maker perspective could therefore omit the lifetime costs and utilities and focus on the cost-saving associated with PTBs averted. Because the purpose of this thesis was to explore the dynamic EVII model in a case study, this scenario analysis does not form part of this thesis.

7.2.2 Deciding on the most appropriate type of decision model

The model used in this analysis was a decision tree model. This is usually appropriate when recurring events are not important (Gray et al., 2011). Brennan et al. (2006) stated in their taxonomy of model structures that a decision tree is an appropriate model choice when interaction between individuals is not needed; the model does not need to be timed; and a cohort model reflects the decision problem accurately. The present decision problem justified the model choice because: 1. interactions between individuals in the model were not likely. 2. Timing would not affect model outcomes, as the time at which the intervention is

used and at which future health outcomes are determined is limited to one year, and health outcomes can be extrapolated for a life-time horizon without the need for accounting for any recursive health states. 3. Individual patient characteristics are not foreseen to change model outcomes significantly.

I developed the decision tree for this decision problem by following the steps proposed by Gray et al. (2011):

a) Tree structure

A decision tree model has a square decision node which indicates a decision point between alternative options (Briggs et al., 2006). A circular chance node shows a point where two or more alternative events are possible for a patient. Each chance node is associated with a probability that shows the likelihood of a particular event occurring at that node. Patients then follow pathways through the decision tree, which are mutually exclusive sequences of events, each associated with a probability of occurring. To the very left of the decision tree, the first probabilities show the probability of an event, but moving to the right, these probabilities become conditional on the previous events.

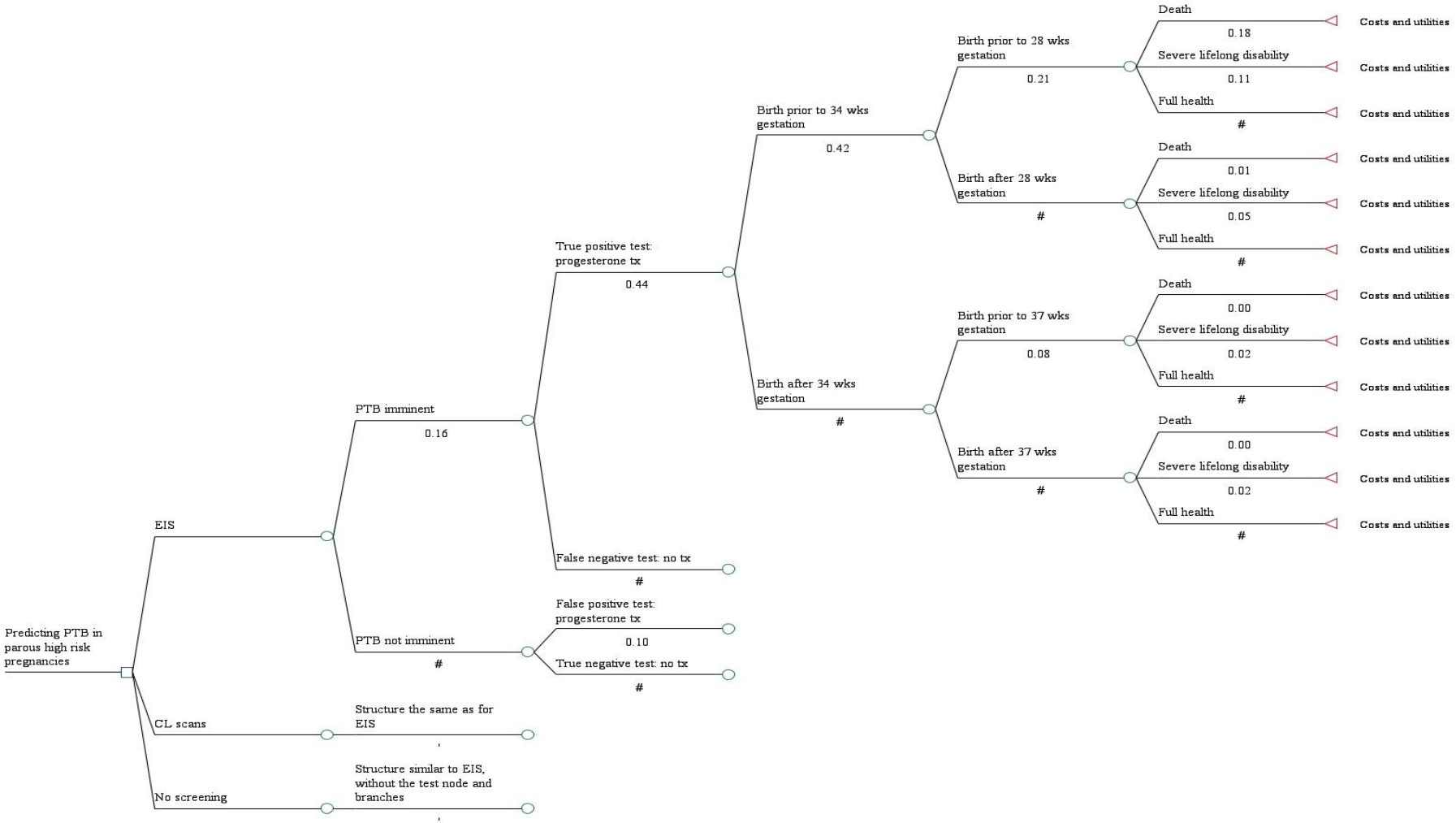
The PTB decision tree model (Figure 7.1) starts out with parous pregnant women that are deemed to be at high risk of PTB based on their history of previous PTB. This includes all pregnant women that have had a previous pregnancy of at least 20 weeks' gestation. In a proportion of these women, preterm birth as defined as birth before 34 weeks' gestation is prevalent, if the woman remains untreated. This is determined by the prevalence in all pregnant women and the sensitivity of previous history as a predictor for PTB. I assume that the history of previous PTB will be known for each of these women, which should be a reasonably realistic assumption in the English NHS. If a woman has a history of PTB, she is sent for EIS screening; or a CL scan or no screening in the comparator strategies. Figure 7.1 only presents one arm of the decision tree, only for EIS screening, for reasons of space.

If a woman tested positive, she would be recommended vaginal progesterone in the form of pessaries for the duration of up to 34 weeks of gestation. With progesterone treatment, the relative risk of having a PTB is reduced. That means that out of the women who were going to have a PTB, a proportion will not have a premature delivery because of the treatment. This results in two branches coming out of that node – the women who give birth prior to 34 weeks of gestation and those who give birth after that.

Costs and health outcomes associated with PTB are significantly influenced by the length of gestation (Boyle et al., 2012). I therefore adopted the approach by Werner et al.

(2011) and divided the length of gestation up further. Among the women who give birth prior to 34 weeks' gestation, a proportion would give birth prior to 28 weeks, and the remainder between 28 and 34 weeks. Among those who give birth after 34 weeks' gestation, a proportion would go on until more than 37 weeks and the remainder would give birth between 34 and 37 weeks. At each different length of gestation, there are three different health states for the baby: full health, severe and lifelong neurological disability and death. Each length of gestation is also associated with a cost of neonatal and maternal care.

Figure 7.1 Illustration of EIS decision tree structure



b) Order of events in the decision tree

For screening tests, there are two possible ways of ordering events (Gray et al., 2011). One is starting with the prevalence of the disease and continuing with the patients that test positive. For this, one needs the disease prevalence and values for sensitivity and specificity. The other way, which may be more intuitive for clinicians but is usually more difficult to populate (Gray et al., 2011), starts out with screening and dividing patients into those who test positive and those who test negative and then continue with those who have the disease. For the latter approach, the positive and negative predictive values are required, and the prevalence of the disease in those who tested positive and negative. The availability of data thus often influences the approach for ordering the events in the decision tree. I followed the more commonly used approach of starting with the prevalence of the disease and let this be followed by the sensitivity and specificity of the respective test because I lacked the data to populate the alternative model for CL scans.

To obtain the number of women with a history of preterm birth to start the decision tree, I used the prevalence of PTB and the sensitivity and specificity of a PTB history, which was obtained from Honest et al. (2009). This resulted in a new estimate for the prevalence of PTB in all high risk pregnant women. EIS, or its comparator interventions CL scans or No screening, would then be used on all the women who had a history of PTB, whether PTB was imminent or not. The remainder of the tree after the different screening interventions was straight forward. Women who tested positive with EIS or CL scans would undergo treatment with progesterone and that would reduce their chance of having a PTB.

c) Estimating probabilities

Each chance node is associated with a probability. As mentioned above, the probabilities after the first chance node are conditional probabilities. With pathways being mutually exclusive, all probabilities coming out of one chance node must sum up to 1. The end of a pathway is therefore associated with a joint probability of all the events in one branch occurring.

d) Payoffs

The term ‘payoffs’ describes costs and utilities that are associated with each final node. Chance nodes along the pathway can also be associated with costs. These costs are summed up to yield the total cost associated with the strategy.

7.2.3 Identifying the evidence and populating the model

Evidence to populate the model was obtained from different sources, mainly the previously described studies (Honest et al., 2009, Werner et al., 2011), but also from other sources, as shown in Table 7.1. For some parameters, obtaining evidence proved challenging and the assumptions made to address this and the impact on model outcomes are discussed below.

There was no data on the number of parous pregnant women and this number was therefore estimated using the number of births per year in England (The Health and Social Care Information Centre, 2012a) and the proportion of parous pregnant women found in a trial on the relative treatment effect of progesterone versus no treatment in preventing PTB by Fonseca et al. (2007). While the data on EIS and CL scans (Honest et al., 2009, Werner et al., 2011) was for singleton pregnancies and this is what is examined in this model, the number of births includes multiple pregnancies. The bias this may cause is fairly small, as over- or underestimation of the numbers of patients in the model would affect only the cost per screening with EIS, and only in very small orders of magnitude.

A preterm birth, as mentioned above, is defined as a birth before 34 weeks of gestation. If the baby is born before 28 weeks’ gestation, the risk of the baby dying or having a severe disability is even higher. The prevalence of preterm births (that is, prior to 34 weeks of gestation) in pregnant women was taken from Honest et al. (2009). A word of caution applies here: an estimate of the prevalence of preterm births in parous pregnant women was not available and I therefore assumed that this would be the same as in all pregnant women.

The likelihood ratios for previous history as an indicator for PTB were obtained from Honest et al. (2009) as were the likelihood ratios associated with CL scans, which the authors obtained by performing an evidence synthesis on relevant trials. As an explanatory note, the positive likelihood ratio is the ratio of sensitivity over (1-specificity) (or alternatively the probability of positive test conditional on prevalent disease over probability of positive test conditional on no disease), and the negative likelihood ratio is the ratio of (1-sensitivity) over specificity (or alternatively, the probability of negative test conditional on disease over probability of negative test conditional on no disease). The greater the positive

likelihood ration the better (e.g. greater than 10 is considered good) and the smaller the negative likelihood ratio, the better (e.g. negative likelihood ratios are between 0 and 1, but the closer to 0, the better). The threshold for a positive test result with CL scans was 25mm, which was described as the most common threshold evaluated in asymptomatic women at 20 weeks of gestation (Honest et al., 2009): a shorter cervix than that would indicate that there was an increased risk of having a PTB. A word of caution for the latter: these likelihood ratios applied to screening of all pregnant women regardless of their risk factors and they may over- or underestimate the likelihood ratios in the population of high risk women based on their history of previous PTBs. To be used in this model, likelihood ratios were converted to sensitivity and specificity.

Sensitivity and specificity of EIS were obtained from a pilot study that had not been published and was provided to me via personal communication with the developers of EIS (Anumba, 2013). These data were preliminary and will be updated once the ongoing trial has reported. Because the pilot study was small, I generated posterior distributions for the sensitivity and specificity parameters by assuming a binomial likelihood for the sample 'number of events' and by using a reference conjugate Beta(1,1) prior distribution.

The relative risk of having a PTB before 34 weeks' gestation when treated with progesterone compared with no treatment was approximately 0.55 (95% credible interval: 0.34, 0.88) in singleton pregnancies, as reported in a trial by Fonseca et al. (2007). This was equivalent to women treated with progesterone having a probability of PTB of 18.1%, conditional on a positive CL scan. The caveat of these data was that the identification of women with potential PTB was based on cervical length shorter than 15mm, that is different from the 25mm used for the predictive ability in the Honest et al. (2009) study. This trial also did not undertake any pre-screening using previous history – and thus represents a sample from a different population of women than what is used in the present model. Furthermore, the lack of data forced me to make the assumption that the relative risk was similar in women who underwent screening with EIS. A Bayesian reference prior was used to account for this added uncertainty, resulting in a probability of PTB when treated with progesterone of 18.6%. The implications for the uncertainty surrounding this parameter are discussed in Section 7.2.4. I was not aware of any negative effects of progesterone treatment on utilities.

There was some evidence from a Chochrane review I decided not to use on the relevative risk associated with progesterone treatment against no treatment in women who had a history of PTB of 0.31 [95% CI 0.14 to 0.69] based on 11 studies (Dodd et al., 2013) and women with a short cervix of 0.64 [95% CI 0.45 to 0.95] based on two studies. The

former was not of use for this model, because treatment of women with history only does not occur in practice. The latter included the study by Fonseca et al. (2007) and another study, but did not report the probability of PTB with no treatment. I therefore used the Fonseca et al. (2007) estimates because the detail provided in the article enabled me to calculate the probability of PTB with progesterone treatment. The implication is that the real relative risk of progesterone against no treatment may be underestimated here, causing potential bias against the no screening, no treatment strategy.

The probabilities associated with the length of gestation and the health states have all been informed by the Werner et al. (2011) study, but were originally reported by Clements et al. (2007) and Moster et al. (2008). The probability of severe disability was an estimate of probabilities of the most prevalent disabilities grouped together. The probability of delivery prior to 28 weeks, which was conditional on a delivery prior to 34 weeks and the probability of delivery prior to 37 weeks conditional on delivery after 34 weeks were obtained from Fonseca et al. (2007). The underlying assumption is that these probabilities are not changed by treatment with progesterone.

The cost associated with each EIS scan was indicative because the price had not been set yet. I assumed costs of £15 in consumables in every patient plus £1 training and maintenance costs per scan and I assumed it would take a consultant 15 minutes at approximately £9.40 to perform the test. The £9.40 estimate was based on inflated consultant costing in the Honest et al. (2009) study. The capital cost per device was assumed at £5,000. The capital cost was annuitised over an assumed time horizon of five years, with the annuity factor calculated using $\frac{1-(1+r)^{-n}}{r}$, where r is the discount rate and n is the number of years the technology will be in use (Drummond et al., 2005). Based on participants' statements in the qualitative study, it was assumed that two devices would be purchased per each obstetric unit, the thinking behind which was that there should be at least one spare unit in case one goes to maintenance.

With that information, the number of scans per year per device could be calculated using the number of parous pregnancies that were determined at risk based on the women's history of PTB, which would be approximately 27,000 women in England per year. The number of obstetric units in England was 231 (UNICEF, 2013), which led to a total potential number of EIS devices in hospitals of 462 and 59 scans per device per year, based on having each woman screened once in their pregnancy. The annuitised price (which was £1,107 based on a purchase price of £5000 for five years) was then divided over the number of scans per year to yield the capital cost per scan, to which training, maintenance and manpower were added to yield the cost per scan of £44 (Table 7.1). This is slightly higher than the cost per scan for

a similar device (Electrical Impedance Spectroscopy for use in cervical cancer screening) which could reflect the further development that has gone into the device.

Where costs were obtained from sources that reported them in US Dollars, they were converted to GBP by using health-care specific purchasing power parities by the World Health Organisation (2015). Costs were inflated to 2014 costs using the Personal Social Services Research Unit (PSSRU) inflation tables (Personal Social Services Research Unit, 2015).

The maternal costs associated with early delivery were reported by Werner et al. (2011) but originally taken from Gilbert et al. (2003). Neonatal costs associated with PTB were informed by Honest et al. (2009) who obtained their estimates by calculating the cost at different lengths of gestation based on the cost associated with different birth weights reported by Petrou (2003), assigning average birth weight to lengths of gestation.

Werner et al. (2011) used costs associated with early intervention in the first three years of a child's life. These included costs of special education, which were not relevant for the payer perspective adopted here. Instead, I used costs reported by Petrou and Khan (2012) that were only of medical nature and estimated for the first 18 years of the babies' lives using a Markov model. The evidence used by the authors stemmed from various countries, including the UK. If a societal perspective were to be adopted, special education costs as well as potential productivity losses could be incorporated in the model (Werner et al., 2011) but this does not apply here.

I also explored medical care cost data of children with low birth weight that was available from a UK study (Stevenson et al., 1996), which could be used by converting the duration of gestation into average birthweight, as was done by Honest et al. (2009). These costs were, however, reported in 1979 values and stemmed from data collection prior to 1980. Major changes in resourcing and management may have occurred since then that would have an impact on these costs, thus making them less applicable than the above costs reported by Petrou and Khan (2012). If the Stevenson et al. (1996) costs were used and inflated to today's value, they exceeded those reported by Petrou and Khan (2012) significantly, leading to potential bias in favour of EIS and CL scans.

The cost for severe disability per year was based on the medical costs associated with cerebral palsy, one of the most common disabilities in premature babies (Werner et al., 2011, US Department of Health and Human Services and Centers for Disease Control and Prevention, 2004).

The cost of single CL scans was taken from the review by Honest et al. (2009), in which it was assumed that one scan would take up to ten minutes of the time of a consultant and one hour of the time of a lab technician doing the analysis. The purchasing cost of ultrasound machines was assumed to be zero, as ultrasound scans with the purpose of preterm birth screening are performed on existing machines in the NHS, in line with the analysis in Honest et al. (2009). If this analysis were to be carried out for other countries, the purchasing cost of ultrasound machines might have to be factored into the analysis. The cost of taking the history of a pregnant woman was assumed to be zero, as this is part of the standard routine assessment of each pregnant woman.

The cost of progesterone injections was reported by Honest et al. (2009) to be at £923.55 for 15 injections. It was anticipated that progesterone would be administered via pessaries following screening with EIS but, for the lack of other cost data, I used the inflated costs reported by Honest et al. (2009) for injections.

Quality-of-life values associated with the three possible health states were based on Werner et al. (2011). The QALY value for severe disability was taken from Odibo et al. (2006) and included the conditions 'cerebral palsy, mental retardation, blindness, deafness and epilepsy' (Werner et al., 2011) that were identified as the most prevalent conditions following a preterm birth by Moster et al. (2008).

Costs associated with severe disability and quality-of-life data were extrapolated over an expected time horizon of 76 years (Werner et al., 2011) and discounted at 3.5% according to the NICE reference case (NICE, 2013), assuming that life expectancy would be the same for babies with and without disabilities. This is a simplifying assumption that may make the cost and the utilities of disability look larger than they actually are. When extrapolated over the lifetime horizon, utility values for both full health and disability were adjusted by population norm utility values (Hawthorme and Osborne, 2005) by multiplication.

Table 7.1 Data inputs used in PTB model

Parameters	Mean values	Reference
Number of births in England	668,936	NHS maternity statistics report (England) (2012)
Proportion of parous women (given birth at least once)	0.44	Fonseca et al. (2007)
Prevalence delivery before 34wks	0.04	Honest et al. (2009)
LR+ of History	4.620	Honest et al. (2009)
LR- of History	0.68	Honest et al. (2009)
EIS sensitivity	0.44	EIS pilot data
EIS specificity	0.9	EIS pilot data
LR+ CL scan (sensitivity)	13.38 (0.21)	Honest et al. (2009)
LR- CL scan (specificity)	0.8 (0.98)	Honest et al. (2009)
Prob. of PTB with progesterone after CL scan	0.181	Fonseca et al. (2007)
Prob. of PTB with progesterone after EIS test	0.186	Assumption
Prob. of PTB with no treatment	0.3214	Fonseca et al. (2007)
Cond. prob. of birth before 28 wks	0.206	Werner et al. (2011)
Cond. Prob. of birth after 37 wks	0.084	Werner et al. (2011)
Prob. disability before 28 wks	0.106	Werner et al. (2011)
Prob. disability 28-34 wks	0.05	Werner et al. (2011)
Prob. disability 34-37 wks	0.024	Werner et al. (2011)
Prob. disability after 37 wks	0.017	Werner et al. (2011)
Prob. death before 28 wks	0.179	Werner et al. (2011)
Prob. death 28-34 wks	0.009	Werner et al. (2011)
Prob. death 34-37 wks	0.002	Werner et al. (2011)
Prob. death after 37 wks	0.0007	Werner et al. (2011)
Cost per EIS screening	£ 44	Based on assumptions
Cost per CL scan	£ 14.38	Honest et al. (2009)
Cost vaginal progesterone	£ 1,115	Honest et al. (2009)
Cost maternal care if delivery at <28 wks	$10,953	Werner et al. (2011)
Cost maternal care if delivery btw. 28-34 wks	$8,153	Werner et al. (2011)
Cost maternal care if delivery at 34-37 wks	$4,627	Werner et al. (2011)
Cost maternal care if delivery at >37 wks	$3,577	Werner et al. (2011)
Cost neonatal care if delivery at <28 wks		Honest et al. (2009)
Cost neonatal care if delivery btw 28-34 wks	£ 19,618	Honest et al. (2009)
Cost neonatal care if delivery at 34-37 wks	£ 15,137	Honest et al. (2009)
Cost neonatal care if delivery at >37 wks	£ 0	Honest et al. (2009)
Cost medical intervention if delivery at <28 wks	£ 123,213	Petrou and Khan (2012)
Cost medical intervention if delivery betw 28-34wks	£ 135,868	Petrou and Khan (2012)
Cost medical intervention if delivery at 34-37 wks	£ 22,829	Petrou and Khan (2012)
Cost medical intervention if delivery at >37 wks	£ 9,700	Petrou and Khan (2012)
Cost severe disability (per year)	$1,175	US Department of Health and Human Services (2004)
Utility death (per year)	0	Werner et al. (2011)

Utility severe neurological disability (per year, these are adj. by pop. norms)	0.61	Werner et al. (2011)
Utility Health (per year, these are adj. by pop. norms)	1	Werner et al. (2011)
Life expectancy (years)	76	Werner et al. (2011)

7.2.4 Synthesizing evidence

For the purposes of this case study, I used available estimates of the required parameters that were obtained mostly from previous economic evaluations or trials as described above. This may mean that relevant evidence was missed. The potential lack of precision introduced by this will not impair this study that seeks to explore the effects of implementation dynamics on economic evaluation outcomes. For the final health economic model on the use of EIS for PTB screening, the model will be updated and it was therefore important to describe the model in sufficient detail to enable the updating of the model at a later stage. This is what I have attempted in Sections 7.2.2 and 7.2.3 and I will continue to describe the handling of uncertainty and simplifying assumptions transparently in this section.

To reflect the uncertainty about the input parameters, I performed probabilistic sensitivity analysis (PSA) varying the uncertain parameter values over a number of 10,000 simulations. I chose beta distributions for all probabilities and QALYs and gamma distributions for all costs, based on guidance by Briggs et al. (2006). I also used the log-normal distribution for likelihood ratios. The probability distributions used in the PSA and shown in Table 7.2 were generated using the data sources shown in Table 7.1, with a few exceptions. All estimates obtained from Werner et al. (2011) were reported with ranges. I assumed that these ranges were 95% credible interval estimates although Werner et al. (2011) did not go into that detail.

The sensitivity and specificity data for EIS were taken from the classification table of the pilot study. I used beta distributions to model each. However, because sensitivity and specificity measures are typically correlated, I also explored the use of a bivariate normal distribution on the log scale. Knowing that there is some correlation, of which the magnitude is currently unknown, I expressed my belief about the correlation (as was done by O'Hagan et al. (2001)) by assigning a covariance to the two measures that I assumed to lie between the individual variances in magnitude, resulting in $m = \begin{pmatrix} 0.44 \\ 0.9 \end{pmatrix}, V = \begin{pmatrix} 0.0247 & 0.01 \\ 0.01 & 0.00309 \end{pmatrix}$. The resulting correlation coefficient then allowed the calculation of the joint and the conditional distributions of sensitivity and specificity. Using the bivariate lognormal

distribution on the log scale led to the marginal distributions of sensitivity and specificity to exhibit smaller variation compared with the separate simulation of these parameters using the beta distribution. For this reason, I continued using the separate simulation of sensitivity and specificity parameters for this illustrative example.

As was done in the Honest et al. (2009) study, I used log-normal distributions for likelihood ratios of CL scans, restricted to a minimum of 1.0001 for the positive likelihood ratio and to a maximum of 0.9999 for the negative likelihood ratio. It needs to be mentioned that the estimates obtained from Honest et al. (2009) reflect a slightly different population, the non-high risk population. I assumed that the diagnostic accuracy may be similar to a high risk population but reflected the greater uncertainty in widened credible intervals.

Instead of using the relative risk estimates in the PSA, I used the probabilities of giving PTB with and without progesterone treatment. This prevented the problem of yielding relative risk ratios greater than one in a small number of samples. The evidence used to generate the posterior distribution for the probabilities of PTB with and without progesterone treatment was obtained directly from Fonseca et al. (2007). To account for uncertainty associated with this estimate, I used a Beta(1,1) reference prior distribution to generate the posterior distribution for relative effectiveness of progesterone treatment after EIS screening.

For lack of data, the credible intervals surrounding the capital costs of purchasing EIS were based on my assumption.

Table 7.2 Uncertain parameters varied in the PSA

Parameters	Mean estimates	95% Credible Interval	Distribution
EIS sensitivity	0.44	(0.16, 0.76)	Beta
EIS specificity	0.9	(0.76, 0.98)	Beta
LR+ CL scan	13.38	(3, 26)	Log-normal
LR- CL scan	0.8	(0.65, 0.9)	Log-normal
Probability of PTB with progesterone after CL scan	0.181	(0.117,0.256)	Beta
Probability of PTB with progesterone after EIS test	0.186	(0.121,0.261)	Beta
Prob disability bef28wks	0.106	(0.091, 0.171)	Beta
Prob disability 28-34wks	0.05	(0.033, 0.097)	Beta
Prob disability 34-37wks	0.024	(0.021, 0.026)	Beta
Prob disability aft37wks	0.017	(0.015, 0.018)	Beta
Prob death bef28wks	0.179	(0.08, 0.493)	Beta
Prob death 28-34wks	0.009	(0.002, 0.086)	Beta
Prob death 34-37wks	0.002	(0.001, 0.004)	Beta
Prob death aft37wks	0.0007	(0.0005, 0.0009)	Beta
EIS cost	£5,000	(£3,000, £15,000)	Gamma
Cost Maternal care if delivery at <28wks	£7,345	(£1,341, £10,728)	Gamma
Cost Maternal care if delivery 28-34wks	£5,467	(£670, £8,046)	Gamma
Cost Maternal care if delivery at 34-37wks	£3,103	(£536, £4,694)	Gamma
Cost Maternal care if delivery at >37wks	£2,399	(£469, £4,023)	Gamma
Cost Neonatal care if delivery at <28wks	£19,618	(£12,505, £50,022)	Gamma
Cost Neonatal care if delivery 28-34wks	£19,618	(£12,505, £50,022)	Gamma
Cost Neonatal care if delivery at 34-37wks	£15,137	(£6,253, £37,516)	Gamma
Cost Medical intervention if delivery at <28wks	£123,213	(£50,022, £150,065)	Gamma
Cost Medical intervention if delivery 28-34wks	£135,868	(£62,527, £187,581)	Gamma
Cost Medical intervention if delivery at 34-37wks	£22,829	(£3,752, £37,516)	Gamma
Cost Medical intervention if delivery at >37wks	£9,700	(£2,501, £18,758)	Gamma
Cost Severe disability (per year)	£941	(£401, £4,006)	Gamma
QALY Severe Neurological Disability (per year)	0.61	(0.5, 0.8)	Beta

7.2.5 Analysing the model

The process of evaluating model outcomes is described by 'folding back' the decision tree. By folding it back, the expected values for each strategy can be calculated. Probabilistic analysis was performed by sampling from probability distributions in Table 7.2. I generated the posterior distribution for incremental costs and benefits, which is commonly represented as a cost-effectiveness plane (Drummond et al., 2005) and presented the probability of the different technologies being cost-effective in the cost-effectiveness acceptability curve (CEAC).

7.2.6 Model evaluation

Model evaluation should be centred around the assumptions and the model structure, the input parameters, distributions that reflect uncertainty, and the output and conclusions (Gray et al., 2011). There are three key processes that aid model evaluation (Gray et al., 2011) that I attempted to follow:

1. Face or descriptive validity, which can be established if assumptions, the model structure and results can be described intuitively: I attempted an intuitive description of the model structure in Section 7.2.2 and described assumptions in the sections on data inputs (7.2.3) and on handling uncertainty (7.2.7).

2. Internal validation and calibration, which tells us whether model inputs relate to outputs with a certain logic. To test this, extreme values and null values can be put into the model to explore whether they would have the expected effects on model results. I used null values for all costs but the costs associated with EIS, CL scans and progesterone and checked whether results and parameters related to each other logically by doing a few PSA runs and checking the relationship between all the PSA run specific parameter values and the resulting costs and QALYs. I also assumed the same distributions for the predictive ability of EIS and CL scans at (at the distribution of the likelihood ratios of CL scans) and re-ran the PSA with a small number of simulations to see whether in each simulation, the resulting costs and QALYs related to the parameter inputs logically.

3. External validity and consistency can be shown if findings can be generalised beyond the evidence used for the model and can be demonstrated to apply in the real world. This is beyond the scope of this thesis.

7.2.7 Handling uncertainty

In addition to presenting parameter uncertainty in the cost-effectiveness plane and CEAC, I also calculated the EVPI, which represents the expected opportunity loss of decision uncertainty. Uncertainty in the output can be broken down into those parameters that contribute the most to it by an EVPPI analysis.

Structural uncertainty reflects assumptions made in building the model. Because this model is immature and needs updating at a later stage, I did not undertake any formal methods of addressing structural uncertainty, such as the model discrepancy method (Strong et al., 2012) or model averaging (Bojke et al., 2009). Instead, I summarise all the assumptions in this model in the list below:

1. The number of births used includes multiple pregnancies but the analysis applies to singleton pregnancies.
2. The incidence of PTB is the same across parous and nulliparous women.
3. The likelihood ratios associated with CL scans in the low risk pregnant population are the same as the likelihood ratios associated with CL scans in the high risk population used here.
4. The relative treatment effect of progesterone compared to no treatment after screening with EIS and after screening with CL scans are approximately the same, apart from adding a reference beta distribution.
5. The relative treatment effect of progesterone compared to no treatment is the same regardless of whether pre-screening with previous history was performed or not.
6. Only a single CL scan is conducted rather than a series of scans. This may have implications in terms of test accuracy and cost.
7. Patients are only screened once using EIS, rather than being exposed to a series of tests. This may have implications in terms of test accuracy and cost.
8. The cost of progesterone injections was used even though pessaries might be recommended in practice.
9. The probability of giving birth prior to 28 weeks conditional on the probability of giving birth prior to 34 weeks was assumed to be independent of changes to the probability of giving birth prior to 34 weeks due to treatment. The same applied to the probability of giving birth prior to 37 weeks conditional on the probability of giving birth posterior to 34 weeks.
10. No difference in life expectancy was assumed for patients in full health and those with any disability.

11. CL scans were assumed not to have any capital costs as ultrasound machines are mainly used for many other purposes and have a long lifetime. This may introduce a slight bias in favour of CL scans.
12. I assumed that each obstetric unit in the country would purchase two EIS devices.
13. Health outcomes were crudely modelled: different conditions were grouped and the same probability, QALY value and cost assumed for all of them.
14. Estimates for likelihood ratios for CL scans were based on a cut-off of 25 mm but the estimate for the relative risk associated with progesterone treatment versus no treatment was based on a 15 mm cut-off.

7.3 Methods for estimating the dynamic expected value of implementation and information illustrated in the PTB screening model

7.3.1 Calculating the value of implementation measures

In this section, I describe how the extended EVII framework introduced in Chapter 4 was applied in the EIS screening case study. I first present results for a static analysis, in which diffusion was ignored and only time-fixed implementation estimates were used. Then, I present results for the dynamic analysis that took account of diffusion. Static and dynamic estimates of implementation were therefore needed to facilitate comparison of the static and the dynamic method. The additional implementation data requirements affected the estimation of the expected value of implementation measures only (see Chapter 4 Table 4.3 for a summary of the different value measures). The diffusion curves required for the dynamic analysis were obtained from the elicitation study described in Chapter 6. I also derived static estimates from there.

To calculate the expected value of perfect implementation, an estimate of optimal implementation levels was needed. I assumed this to be 462 devices, based on the number of obstetric units in England (231 according to UNICEF (2013)) and assuming that two devices are needed for each obstetric unit (see Chapter 6), resulting in 462 adoptions for England. In the absence of information to suggest otherwise, I assumed that the share of the other two technologies over the remaining adoptions was equal, as was detailed in Section 4.2.4 in

Chapter 4. Table 7.3 shows a summary of the data used for each of the expected value measures.

Calculating the EVPIM also required knowledge of current implementation of the alternative technologies. Current implementation refers to the level of implementation that can be achieved currently, that is, without the performance of any further research or implementation initiatives. For the dynamic EVPIM analysis, the elicited baseline diffusion curve was used (see Table 7.3). For the static analysis, I used the estimate of the maximum attainable number of adoptions from that elicited diffusion curve (the mean estimate was 140). At the time of technology appraisals, implementation of a new technology is commonly zero percent. Using this as the baseline implementation level would, however, over-inflate the EVPIM, as the recommendation decision alone is commonly assumed to trigger a certain level of implementation (this was backed up by findings from the review of diffusion curves in Chapter 3). Of course, the assumption of a sudden jump in implementation is not necessarily realistic, highlighting the limitations of using static analysis.

Table 7.3 Data requirements for estimation of Expected Value measures

Value concept	Implementation data required	Data used in static analysis	Data used in dynamic analysis
EVPIM	Current implementation	140 adoptions (=30.3%) (elicited 'current' maximum attainable adoptions)	Elicited current diffusion curve
	Perfect implementation	100%	100% in each period
EVPIM^{R}	Current implementation	140 adoptions (=30.3%) (elicited 'current' maximum attainable adoptions)	Elicited current diffusion curve
	Implementation at perfect information	200 adoptions (=43.3%) (assumed)	Elicited 'Both research studies' diffusion curve parameters, but with maximum adoptions of 200
EVSIM^{Rpp}	Current implementation	140 adoptions (=30.3%) (elicited 'current' maximum attainable adoptions)	Elicited current diffusion curve
	Implementation at perfect parameter information for EIS + Prog parameters	150 adoptions (=32.5%) (assumed)	Elicited 'Both research studies' diffusion curve parameters, but with maximum adoptions of 150
EVSIM^{S1+S2}	Current implementation	140 adoptions (=30.3%) (elicited 'current' maximum attainable adoptions)	Elicited current diffusion curve
	Implementation with respective research evidence available	142 adoptions for EIS trial and 144 adoptions for Progesterone trial (elicited study specific maximum attainable adoptions)	Elicited study specific diffusion curve

To calculate the EVPIM^{R}, estimates of implementation levels that would occur when all decision uncertainty had been resolved were needed. This was not elicited and I assumed that 200 adoptions (43%) could be achieved by resolving all uncertainty for illustration purposes. The number of 200 out of 462 possible adoptions achievable through research was deliberately low, because the qualitative study reported in Chapter 5 revealed that there were many factors affecting the implementation of EIS, not only research. This finding was confirmed in the elicitation study in which experts pointed out that 'some trusts never adopt new technology' or that the predictive ability of EIS was too low to warrant greater implementation.

Calculation of the respective EVSIM of the two research studies on 1. the treatment effect of progesterone against no treatment after EIS screening and 2. the parameters associated with the predictive ability of EIS (sensitivity and specificity) required quantitative estimates on diffusion for each. These were obtained by applying the elicited proportional gain in implementation for each of the studies κ to each period of the difference between the 'with further research' diffusion curve and the 'without further research' diffusion curve (61% for the relative treatment effect of progesterone versus no treatment after screening with EIS study, and 39% for the EIS predictive ability study). For the static analysis, I used the elicited study-specific maximum attainable number of adoptions, which resulted in an increase of approximately four purchases for the progesterone trial and of approximately two purchases for the EIS screening trial. It is important to note here that the with research diffusion estimates were based on the elicitation exercise in which the effects of possible study results were not considered. The EVSI is therefore calculated assuming that diffusion and expected net benefit are independent of each other, as was shown in Equation (4.12) in Chapter 4. This limitation needs to be addressed in future research. To obtain an intuition of the potential impact on EVSIM results, I explore this issue in Section 7.4.4, where I apply Equation (4.11) from Chapter 4 in an extension to this case study.

For the $EVSIM^{Rpp}$ analysis, knowledge of the level of implementation when uncertainty surrounding these parameters is completely resolved was required. This was not elicited but I assumed it to be at 150 adoptions. This was based on assuming diminishing marginal returns to research on the same parameters and knowledge of the maximum attainable adoptions with both research studies, and without further research.

7.3.2 Calculating the value of information measures

The calculation of EVPI and EVPPI are straightforward and methods are described elsewhere (Briggs et al., 2006, Claxton et al., 2012, Strong et al., 2014). The evaluation of a specific research study using the EVSI, however, deserves a more detailed description. To calculate the EVSI, it was necessary to simulate the research that was planned. The process of trial data simulation has been described previously (Ades et al., 2004, Strong et al., 2015, Brennan and Kharroubi, 2007) and I follow the steps outlined there. The two planned research studies were 1. to investigate the relative treatment effect of progesterone compared with no treatment after screening with EIS, and 2. to study the predictive ability of EIS in identifying women that would give PTB.

For 1., one possible research design after recommendation of EIS could be an observational study in which progesterone treatment was given to those patients who tested positive. The omission of a control arm could be caused by ethical issues in recruiting patients for the no treatment arm, once screening and treatment are recommended. To simulate such a one-arm observational study, the response to treatment with progesterone after screening with EIS was estimated. A statistical model for the data that would be collected had to be specified, together with the sample size, and then a dataset from the proposed study could be simulated for each 'row' of the PSA on that parameter. The probability of having a PTB after treatment with progesterone was represented by a beta distribution. For the trial simulation, a sample size of 150 patients was assumed. For each row of the PSA, I generated a sample of data $x^{(k)}$ for those 150 patients (where *k* is the number of draws from the PSA), sampling from a $Binomial\ (P_{PTB.P}^{(k)}, 150)$ distribution (where $P_{PTB.P}$ is the probability of having a PTB given treatment with progesterone). Given the simulated data, the implied model parameters for the decision model could be calculated. I calculated the EVSI using the Generalized Additive Model (GAM) regression method developed by Strong et al. (2015), which entailed regressing the incremental net benefits on the simulated data for each run of the PSA and estimating the EVSI using Equation (4.4) in Chapter 4.

I followed the same process for simulating a study of the predictive ability of EIS. For this, I used the PSA output for the sensitivity and specificity of EIS and assumed that the prevalence of 13% PTB was fixed. I assumed that sensitivity and specificity were uncorrelated for simulating the data, as I have also done within the PSA. I simulated a trial with a sample size of 300 non-randomised high risk pregnant women. All pregnant women would be tested for PTB using EIS screening. There are therefore four different possible outcomes: patients could have tested positive and had a PTB (quantified by the sensitivity times the prevalence), patients could have tested negative and had a PTB (1-sensitivity times prevalence), or patients could have tested positive or negative and not had a PTB (1-specificity or specificity times prevalence, respectively). Assuming the prevalence was fixed, I generated data for sensitivity and specificity in independent simulations. This enabled me to know the sample size in each data simulation: for the sensitivity simulation, the sample size was determined by the prevalence in the high risk population (approximately 13% based on prevalence in all pregnant women of 0.04 and number of parous pregnant women with history of PTB as well as sensitivity and specificity of history) times the overall trial population, resulting in $Binomial\ (P_{EIS.Sens}^{(k)}, 39)$. For the specificity, the sample size was the remainder, resulting in $Binomial\ (P_{EIS.Spec}^{(k)}, 261)$.

7.3.3 Accruing the EVII measures over the population and decision relevance horizon

In order to compare the EVII measures with the cost of a research study, the above value measures need to be accrued over the relevant population and decision relevance horizon. The population of women screened annually was approximately 27,000 women. I assumed a decision relevance horizon of five years.

When specific implementation measures or research studies are to be evaluated against their cost, their value will only accrue for the proportion of the population that can be reached with the cost-effective technology, which will be shown to be EIS. This proportion of the population is achieved by adjusting the overall population by the implementation that can be achieved with the implementation measure.

7.3.4 The timing of implementation measures and their effects

To calculate the EVSIM of both research studies with a delay in time, I assumed the study on the predictive ability of EIS would report within two years of the decision, and the progesterone treatment effect study would report within three years of the recommendation decision. I assumed that the reported research evidence would have an effect on implementation right away. In the first years up to the point at which the research reported (two or three years after the decision), implementation would follow the elicited baseline uptake curve and after that it would follow the curves that were elicited for the two research studies. This jump is a simplification because the post-research curves were elicited under the assumption that those research results were available at the start of the implementation process. The bias that may result from this simplification is an over-estimation of the EVSIM, stemming mainly from the first periods after the jump in the implementation curve. Ideally, different implementation curves for different timings at which the research reports would have been elicited to solve this problem.

7.3.5 Assessing the residual EVP with implementation measures

The residual EVP with implementation measures can be calculated by subtracting the EVSIM associated with the planned implementation measure from the EVP. If the

implementation measure is research, it is important to evaluate its value in terms of both, the EVSI and the EVSIM and subtract both from the EVP. For this evaluation, the population values of the EVP aggregated over all periods until the decision relevance horizon should be used to reflect the value of the implementation measure to the payer. The implementation measure is then worth doing if the population EVSIM exceeds its cost.

7.3.6 Practicalities of building the model and the EVII analysis

The cost-effectiveness model was built using MS Excel. The PSA was performed in Excel using VBA. The individual and grouped EVPPI results were obtained using the SAVI online tool (http://savi.shef.ac.uk/SAVI/). Further analysis on the other value measures (for instance, the EVSI, EVPIM, EVSIM) were performed in R.

The EVPPI and EVSI analyses were accomplished using the GAM regression to estimate the EVPPI developed by Strong et al. (2014). EVPPI analysis has so far been a computationally expensive exercise because the analysis required sampling from all uncertain parameters for a certain number of times within an inner loop when the parameter of interest was fixed and sampling again from all uncertain parameters with a new value for the parameter of interest in an outer loop (Briggs et al., 2006). The problem of the computational burden was even larger when calculating the EVSI in cases where the prior distribution was not conjugate to the data likelihood, because a Markov Chain Monte Carlo simulation would be required to obtain the posterior (Strong et al., 2015). With the nonparametric regression-based methods developed by Strong et al. (2014) and Strong et al. (2015), computation times of the EVPPI and EVSI are significantly reduced by avoiding the nested two-level approach.

7.4 Results of the EVII analysis applied in the PTB screening model

7.4.1 Cost-effectiveness results and uncertainty in PTB screening model

Based on the PSA, EIS was expected to be dominating against both CL scans and no screening strategies. The PSA with 10,000 simulations resulted in EIS being the technology with the highest expected net benefit (Table 7.4). All costs and QALYs were scaled down to

a per person level based on all screened persons. Each screening strategy was associated with approximately 22 QALYs – these were discounted QALYs the average baby born would accrue over their lifetimes. EIS screening provided a small QALY gain over CL scans and over no screening. This was caused by the identification of more women going on to give PTB who could then be treated. In some of these treated women, PTB could be averted. For all screened women, this translated into a gain of 2,192 QALYs that EIS provided over CL scans, and a gain of 2,601 QALYs that EIS provided over the no screening strategy.

Table 7.4 PSA results for EIS in PTB model (based on 10,000 simulations), scaled to per person screened

Threshold = £20,000 / QALY	EIS	CL scans	No screening
Expected Costs	£32,649	£35,774	£38,493
Expected QALYs	22.36	22.33	22.30
Expected number of PTBs*	3,279	3,888	4,418
Expected number of detected PTBs*	1,963	939	0
Expected number of PTBs averted*	1,139	529	0
Expected number of false negatives*	2,454	3,478	0
Expected number of false positives*	2,374	365	0
Cost per screening strategy without treatment or consequences*	£1.2m	£0.4m	£0
Cost of consequences:			
disability*	£15.5m	£16.2m	£16.8m
Maternal and neonatal treatment and delivery*	£870m	£958m	£1.03bn
ICER against CL scans	Dominating	-	NA
ICER against no screening	Dominating	Dominating	-
Expected Net Monetary Benefit	£414,481	£410,735	£407,488

* not scaled to per person screened

EIS was associated with expected costs that were smaller than those associated with CL scans and no screening strategies (Table 7.4). For all screened women the expected saving that EIS provided over CL scans was £136.77 million; and the expected saving of EIS over no screening was £161.76 million.

Since EIS was still in development, the maximum price under which EIS would still be cost-effective against its comparators may be of interest to the manufacturer. I therefore performed a threshold analysis and found that the maximum price of EIS, at which EIS was still cost-effective against CL scans at £20,000 per QALY, was £991,507 (rather than £5,000). This was the price per device conditional on the other costs per scan being fixed at: £15 for disposables, £9 for manpower, £1 for maintenance and training. The resulting cost

per scan of capital and other costs would be £3,733. Against no screening, the maximum price would be at £1.87 million, resulting in a maximum permissible cost per scan of £7,002.

The PSA results indicate that the expected values presented in Table 7.4 were associated with uncertainty. This is shown in the cost-effectiveness planes in Figure 7.2 and Figure 7.3 in which part of the distribution of the incremental net benefit when comparing against CL scans lies in the North West (the dominated) quadrant (11%). A very small part of the incremental net benefit distribution lies in the North East quadrant in which the location above or below the threshold diagonal determined whether EIS was cost-effective against CL scans in that particular run of the simulation. This is reflected in the CEAC in Figure 7.5, in which the probability of EIS being cost-effective appears to be slightly lower at low thresholds than at higher thresholds. The largest part of the incremental net benefit distribution of EIS against CL scans lies in the South East (the dominating) quadrant (Figure 7.2). Against no screening, all of the distribution lies in the South East quadrant (Figure 7.3). CL scans are clearly dominating against no screening (Figure 7.4), with only 0.4% of the incremental net benefit distribution lying in the North West quadrant. Figure 7.4 also shows a smaller spread of the incremental net benefit distribution of CL scans against no screening compared with the EIS comparisons, indicating greater certainty for the CL scans versus no screening comparison.

The EVPI associated with this decision is £141.92 per person (see Table 7.5). Accrued over the population of women screened in England (approximately 27,000 women screened), the annual population EVPI is £3.83 million. Over a decision relevance horizon of five years, the population EVPI amounts to £19.16 million.

Figure 7.2 Cost-effectiveness plane of EIS against CL scans

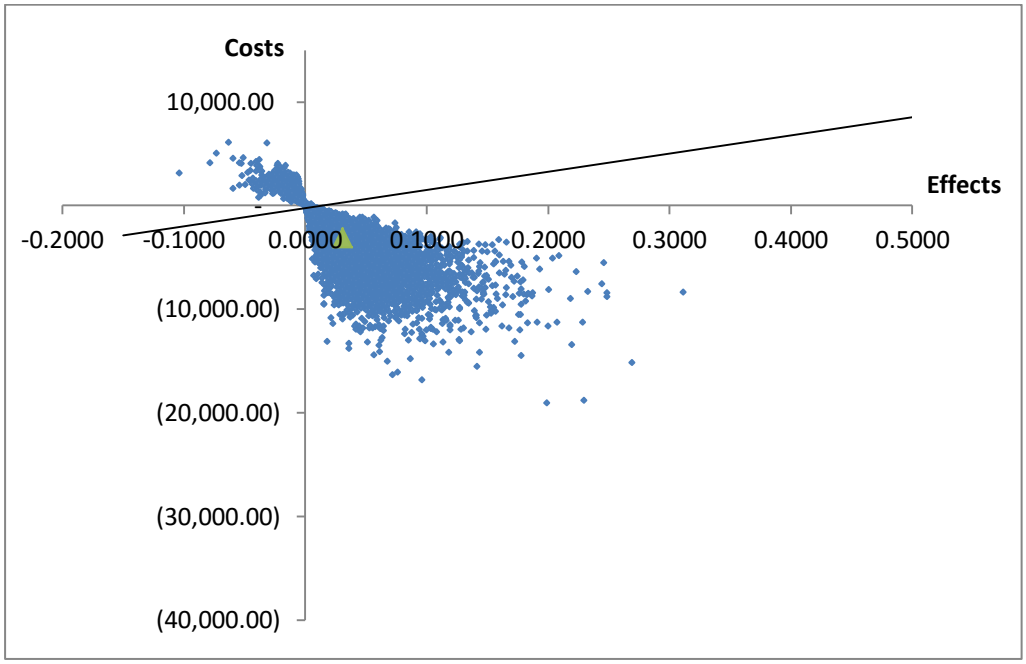


Figure 7.3 Cost-effectiveness plane of EIS against no screening

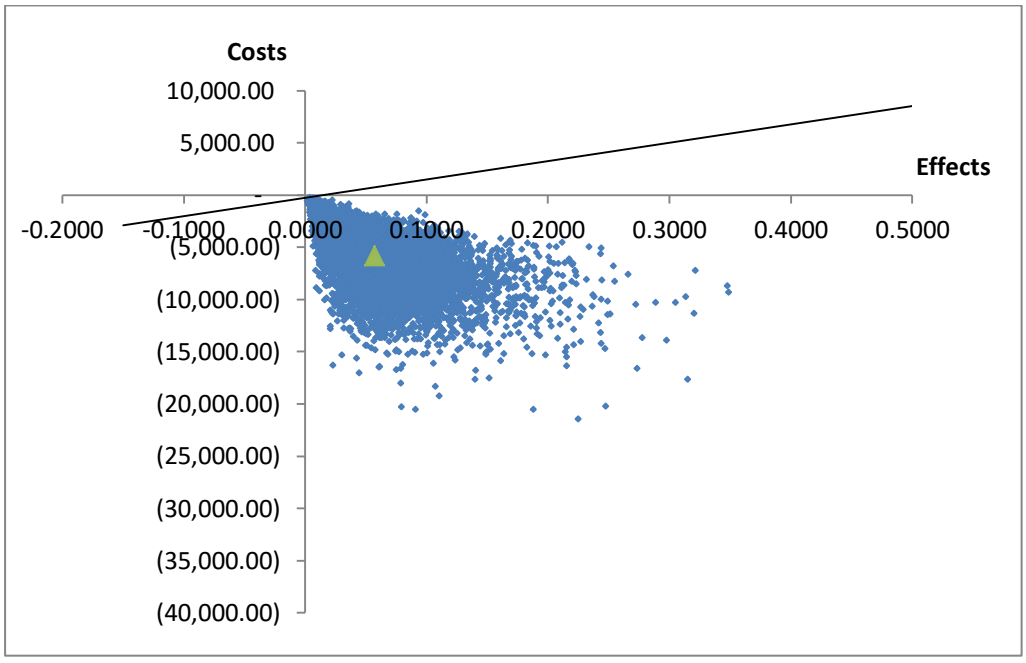


Figure 7.4 Cost-effectiveness plane of CL scans against no screening

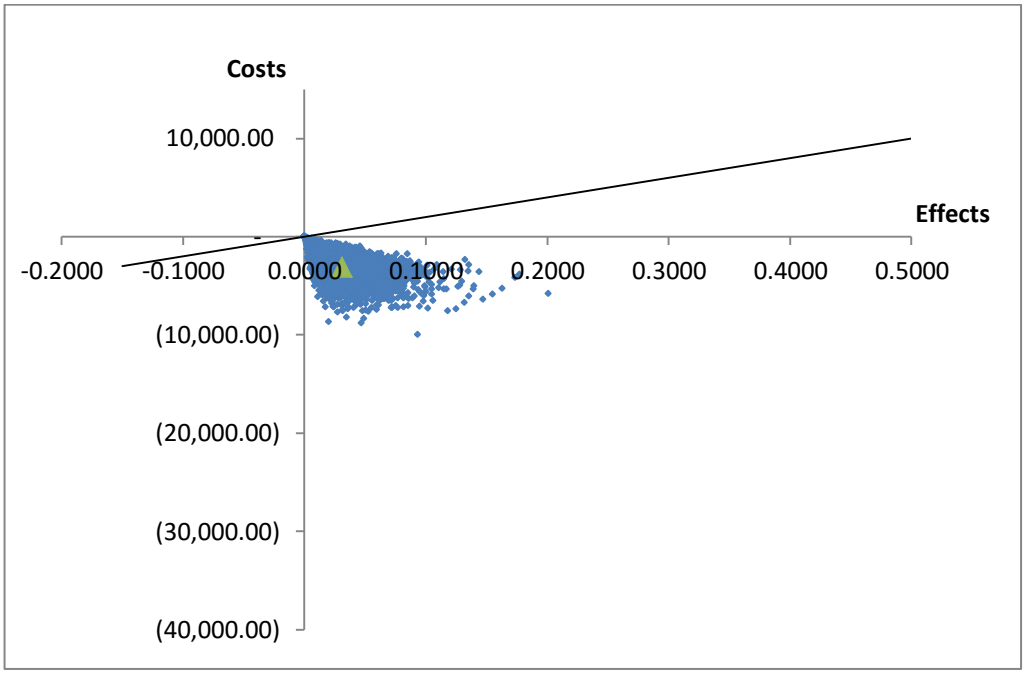
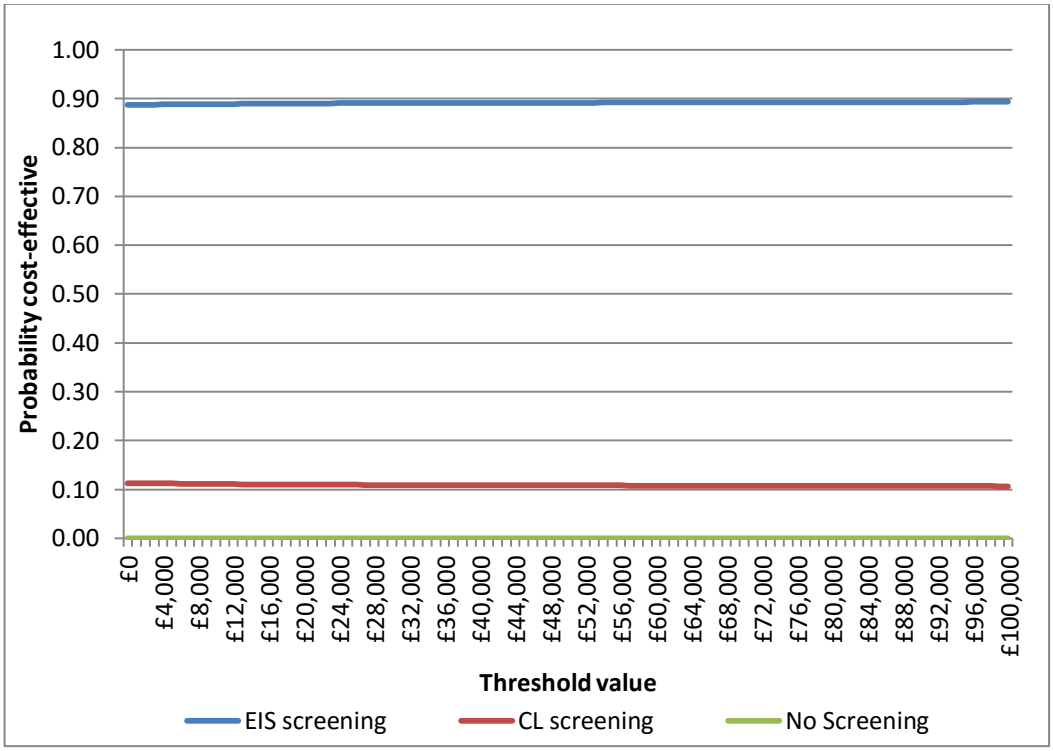


Figure 7.5 Cost-Effectiveness Acceptability Curve for EIS against CL scans and no screening



The reason for EIS being so clearly dominating lay in the fact that its use implied little additional costs (estimated at £44 per person screened, compared with £12 for CL scans and £0 for no screening) and saved some women from having a very costly PTB. The facts that 1. the predictive ability of EIS was not perfect (especially the sensitivity was low at approximately 44%) and that 2. the treatment effect of progesterone compared with no treatment was not very large do not impair this strategy's dominance over no screening, no treatment. This was because one PTB averted could prevent the payer from incurring a cost of up to £150,000 for maternal care, neonatal care and medical intervention in 18 years of the child's life (this is ignoring the potential cost of disability), in the worst case of a PTB prior to 28 weeks of gestation.

The main cause for EIS performing strongly compared with CL ultrasound scans was that the negative likelihood ratio associated with CL scans was very high (at 0.8). This meant that CL scans were not very useful in ruling out that a woman would have a PTB because they produced a large number of false negatives (almost 80% of the women in whom a PTB was imminent, compared to 56% false negatives with EIS, based on the mean estimates from the model). This led to many fewer women treated (940 women out of the ones with PTB imminent with CL scans compared with 1,960 with EIS screening) and fewer avoided cases of PTB as a result (529 compared to 1,139).

To explain the clear-cut results against no screening, one aspect in which EIS could perform worse than no screening was that a woman could falsely be tested positive and treated with progesterone at a cost of approximately £1,100. The cost of even one PTB averted would however be much larger, because, only in terms of neonatal care, maternal care and early intervention and ignoring lifelong costs, the cost-saving per PTB averted would be at approximately £72,000 (when a weighted average for the costs prior and posterior to 28 weeks' and 37 weeks' gestation was taken). In order for no screening to perform better than EIS screening, EIS would therefore have to produce at least 65 False Positives for each PTB averted. On average, however, EIS produced only 2False Positives for each PTB averted (in combination with progesterone treatment).

It is no surprise, given the decision problem, that costs and QALYs appeared to be strongly correlated in the cost-effectiveness planes in Figure 7.2 to Figure 7.4. The correlation was explained by large parts of the costs arising with poor health outcomes for the child.

To summarise this section, EIS for use in PTB screening together with progesterone treatment of women that tested positive was shown to be a dominating strategy against the

strategies of CL scans and progesterone treatment and no screening and no treatment. The results of the cost-effectiveness model were insensitive to the price of EIS. In fact, based on this model, the manufacturer could charge up to almost £1 million per device, given that the other cost parameters were reflected accurately.

7.4.2 The static EVII analysis applied in the PTB screening model

The above analyses showed the per person EVPI to be at approximately £142 per person, or £19 million for the population and a decision relevance horizon of five years (Table 7.5). The EVPPI analysis identified the sensitivity of EIS as the parameter driving the largest part of the EVPI (39%). Other individual parameters causing uncertainty were the relative risk ratio of treating with progesterone compared to not treating the women who tested positive with EIS and the negative likelihood ratio of CL scans, but both contributed less than 1% of the EVPI. I selected a few combinations of parameters based on my beliefs on what may contribute most to decision uncertainty and evaluated the grouped EVPPI. The result was that parameters for the predictive ability of both CL scans and EIS screening together with the treatment effect of progesterone after both types of screening explained approximately 100% of the EVPI.

The burden caused by low implementation was larger than the burden caused by uncertainty, costing the payer £3,742 per patient, which, over the population of women screened and the decision relevance horizon of five years, amounted to £472 million (Table 7.5). A significant part of the EVPIM could be addressed using research. This was revealed by the Research EVPIM, which was large, at £700 per person which translated into £88 million for the affected population in England over a five year decision relevance horizon. This $EVPIM^R$ meant that the value of resolving all decision uncertainty in terms of changing implementation was at £88 million. This suggests that it may be worthwhile to consider research studies as a means of increasing implementation, thus reducing the burden to the payer of financing other sub-optimal screening technologies. It needs to be mentioned, however, that the $EVPIM^R$ was based on assumptions surrounding the level of implementation that could be achieved when all uncertainty had been resolved. The EVSIM is more specific in telling us what the value of an actual research study would be in terms of increasing implementation. The possible reduction in the EVPIM that could be achieved by complete elimination of all parameter uncertainty around the predictive ability of EIS and the progesterone treatment effect had a value of approximately £116 per person or £1.55

million for the affected England population over five years (see the perfect parameter information Research EVSIM in Table 7.5).

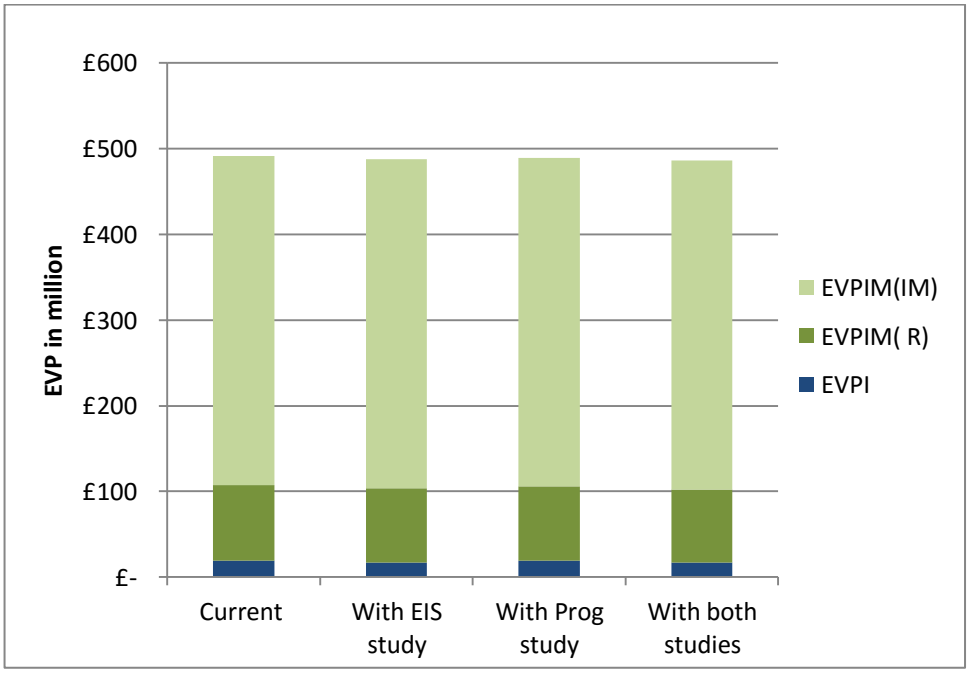
The research study on the progesterone treatment effect after screening with EIS in a two-arm randomised controlled trial with 300 patients recruited had an EVSI of approximately 21p per person. The EVSI of the other research study on the predictive ability of EIS was £52 per person. The EVSIM associated with both studies were £42 and £28 per person, respectively. Note that the EVSI, EVSIM and $EVPIM^R$ population values were adjusted by implementation.

Together, the EVP therefore was £3,884 per person, or £491 million for the affected England population over five years. The Expected Value of Research was £3.23 million for the EIS study and £1.65 million for the progesterone study for the England population over five years. Compared to the EVP, these are small numbers. This is mainly explained by implementation remaining small even with further research evidence becoming available. This has two effects: 1. on the adjusted population estimate that the value measures accrue for and 2. on the implementation-related value measures themselves.

Table 7.5 The static EVII results in the PTB screening model

		Per person	Population per annum	Population over time horizon of 5 years (discounted)
Uncertainty	EVPI	£142	£3.83 million	£19.16 million
	EVPPI (EIS sens+spec)	£58	£1.56 million	£7.3 million
	EVPPI (Tx effect Pg)	£0.32	£8,733	£40,811
	EVSI (EIS)	£52	£428,433	£2.15 million
	EVSI (Pg)	£0.21	£1,577	£8,862
Implementation	EVPIM	£3,742	£101 million	£472 million
	$EVPIM^{R}$	£697	£18.83 million	£88 million
	$EVSIM^{Rpp}$ (EIS + Pg)	£116	£327,922	£1.55 million
	EVSIM (EIS)	£28	£232,127	£1.08 million
	EVSIM (Pg)	£42	£351,126	£1.64 million
Both	EVP	£3,884	£105 million	£491 million
	EVR (EIS)	£80	£660,560	£3.23 million
	EVR (Pg)	£42	£352,703	£1.65 million

Figure 7.6 The EVII analysis chart for the static EIS analysis



The static EVP and possible reductions in it with the two research studies are presented in Figure 7.6. The residual EVPs with research studies were not considerably smaller than the without research EVP because the Research EVPIM was only reduced by between 1-2% for both studies. Of course, it may be worthwhile conducting these research studies as long as the costs of the planned study do not exceed the respective expected value of research, that is, the sum of EVSIM and EVSI. At EVRs of £3 million and £1.65 million for the two research studies respectively, this could be possible. At a crude cost estimate of £2 million for either of the trials, the study on the predictive ability of EIS would provide a positive net gain, while the study on the relative treatment effect would not.

To summarise this section, the large EVPIM together with the relatively smaller EVPI resulted in a large burden of EVP of £491 million for the affected England population over five years. This large burden was caused mainly by two factors: that EIS was clearly better than the comparative strategies, exhibiting a much larger expected net benefit, and that implementation of EIS was expected to be relatively low at approximately 30% throughout its lifetime. The proposed research studies only caused a relatively small reduction in the EVP by the EVR because 1.) the EVPI was small and reductions in it posed a negligible improvement in the overall EVP and 2.) the gain in implementation that could be achieved by those research studies was small, resulting in an improvement of implementation of only 3%. In absolute terms, the EVRs of the respective research studies may still be large enough at £1.65 and £3.23 million to potentially make them worthwhile if their costs fell below the expected value of research. These results indicate that other implementation measures may reduce the large EVPIM further. The static analysis was based on static baseline implementation estimates that were unrealistic. Dynamic analysis is therefore required and will be demonstrated in the next section.

7.4.3 The dynamic EVII analysis applied in the PTB screening model

7.4.3.1 Dynamic analysis when research evidence is available at the time of recommendation

The above analysis ignored the dynamics of implementation that occur with and without the investment in implementation measures and research. In this section, I present the results when these dynamics were considered (Table 7.6) and the new residual EVPs in Figure 7.7. Results of the dynamic analysis only differed in the $EVPIM^{R}$, the $EVSIM^{Rpp}$, the EVPIM, EVP and the EVSIM values but not in the EVPI and EVPPI (Table 7.8).

The dynamic analysis considered the difference in expected net benefits in each period, resulting in different estimates for the implementation-related expected value measures in each period. These differing values are illustrated in Table 7.7. To obtain the expected population values presented in Table 7.6, I multiplied the average value measures over all periods up to the decision relevance horizon (shown in Table 7.7) with the expected patient population. In the case of EVSIM and EVSI the population values were adjusted by the achievable uptake in each period.

The trends of the different EVII measures are explained by the following: The EVPIM is falling in each future period because the gap between perfect implementation and the increasing uptake curve is closing. The Research EVPIM, $EVSIM^{Rpp}$, and the EVSIMs of both studies rise over time as the gap between the with and without research diffusion curves widens.

In this case study, results for the measures influenced by implementation obtained from the dynamic analysis differed from those from the static analysis (Table 7.8). In some cases they were larger (EVPIM, $EVSIM^{Rpp}$ and both EVSIMs) and in one case the dynamic value was smaller ($EVPIM^{R}$). The reason for the differing values is illustrated through the example of the two research studies and their effect on implementation in Figure 7.8. In the static analysis, it was assumed that the recommendation decision was followed by an immediate implementation of the baseline uptake, shown by the dashed grey line. Further research studies on the progesterone treatment effect and the predictive ability of EIS before the research recommendation would result in immediate implementation shown by the solid grey line in Figure 7.8. More realistically, a recommendation decision would trigger the much slower implementation process illustrated by the dashed blue line for no further research and the solid blue line for further research studies having completed (Figure 7.8). It becomes evident from Figure 7.8 that the gap between the blue lines is much larger than between the grey lines, even when the average over all periods is taken. This led to the dynamic analysis exhibiting larger values for the EVSIM associated with the different research strategies. This finding is not generalisable as the difference between dynamic and static results depends on the static before and after implementation values as well as the distance between the dynamic curves in each period. For the $EVPIM^{R}$, for instance, a jump from 140 to 200 adoptions was assumed in the static analysis, leading to a greater difference between the static solid lines than between the dynamic curves.

Figure 7.7 shows that the reduction in the EVP achievable with the designed research studies remained small. This was because the EVP increased. The EVSIM also increased but

the reduction in the EVP that is achievable with the research studies, that is the EVR, remained comparatively small and posed only just above 1% of the EVP.

Table 7.6 The dynamic EVII results in the PTB model

		Per person	Population per annum (average)	Population over time horizon of 5 years (discounted)
Uncertainty	EVPI	£142	£3.83 million	£19.16 million
	EVPPI (EIS sens+spec)	£58	£1.56 million	£7.3 million
	EVPPI (Tx effect Pg)	£0.32	£8,733	£40,811
	EVSI (EIS)	£52	£428,433	£2.15 million
	EVSI (Pg)	£0.21	£1,577	£8,862
Implementation	EVPIM	£4,988	£126 million	£631 million
	$EVPIM^{R}$	£582	£3.30 million	£16.5 million
	$EVSIM^{Rpp}$ (EIS + Pg)	£343	£1.43 million	£7.15 million
	EVSIM (EIS)	£129	£383,811	£1.92 million
	EVSIM (Pg)	£193	£645,862	£3.23 million
Both	EVP	£5,130	£130 million	£650 million
	EVR (EIS)	£181	£812,244	£4.07 million
	EVR (Pg)	£193	£647,439	£3.23 million

Table 7.7 Per period EVII estimates in dynamic analysis

Period	EVPIM	$EVPIM^{R}$	$EVSIM^{Rpp}$	EVSIM EIS trial	EVSIM Prog trial
1	£5,299	£136	£85	£32	£48
2	£5,190	£348	£217	£82	£123
3	£5,032	£609	£374	£142	£212
4	£4,827	£843	£499	£188	£281
5	£4,592	£974	£538	£200	£301
Average	£4,988	£582	£343	£129	£193

Table 7.8 Comparison of static and dynamic analysis in the PTB model

Per person values	Static analysis	Dynamic analysis
EVPIM	£3,742	£4,988
$EVPIM^{R}$	£697	£582
$EVSIM^{Rpp}$ (EIS + PG)	£116	£343
EVSIM (EIS)	£28	£129
EVSIM (Pg)	£42	£193
EVP	£3,884	£5,130
EVR (EIS)	£80	£181
EVR (Pg)	£42	£193

Figure 7.7 The EVII analysis chart for the dynamic analysis

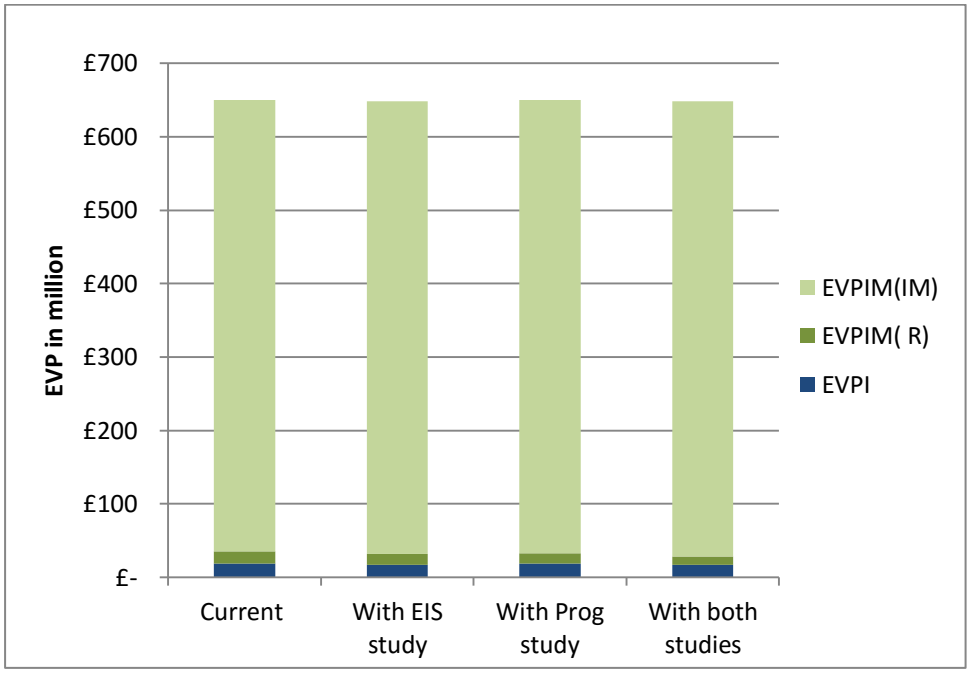
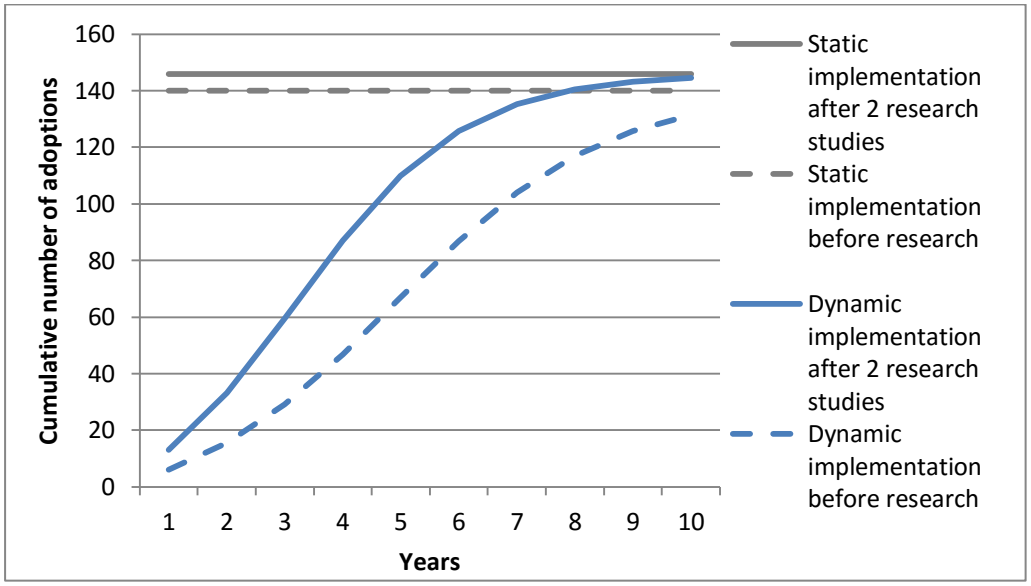


Figure 7.8 Comparison of static and dynamic implementation estimates



7.4.3.2 Dynamic analysis when considering the timing at which research reports

When the timing of research was considered and the results of the research only became available after the recommendation decision, the values of the EVSIM were lower than the values of the EVSIM when research reported before the recommendation decision (Table 7.9) because the payer would not accrue any benefit from it for some years. This is presented in more detail in Table 7.10. The newly calculated EVSIM values, however, were still larger than the values obtained from the static analysis.

The fact that the decrease was considerable could be explained by the relatively large difference between diffusion curves especially in the first few periods and the discounting that would result in greater importance being placed on the periods in the near future. This is shown through the implementation curves for the two studies and the kinks they exhibit at the time of reporting of the results compared with the (dotted) curves that could have been obtained if research reported now, presented in Figure 7.9. This illustrates that the time at which research reports could have a large effect on the EVSIM. For example, if research reported only after the five year decision relevance horizon, the EVSIM would be zero.

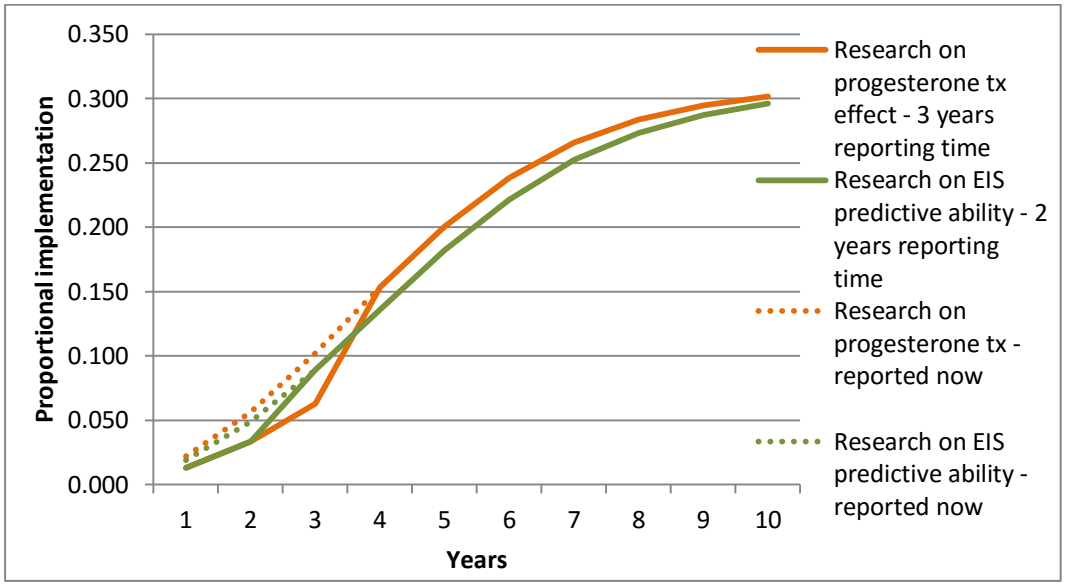
Table 7.9 EVSIM values at future reporting times

Study – time to reporting	Per person	Population per annum	Population over decision relevance horizon of 5 years (discounted)
EVSIM (EIS) – 2 yrs	£106	£33,798	£168,992
EVSIM (Pg) – 3 yrs	£116	£36,626	£183,132

Table 7.10 Per period and person EVSIM estimates in dynamic analysis

Period	EVSIM EIS trial with 2 years reporting time	EVSIM Prog trial with 3 years reporting time
1	£0	£0
2	£0	£0
3	£141.50	£0
4	£187.57	£281.35
5	£200.42	£300.63
Average	£106	£116

Figure 7.9 Implementation curves with different research studies and reporting times



In summary, using dynamic implementation estimates, as opposed to their static counterparts, had a significant effect on all EVII measures, which in this example were mainly larger with implementation dynamics than with a static analysis. It is worth noting that dynamic analysis does not always generate larger or smaller results than the static analysis – the sign of the change depends on the assumptions made about the static implementation levels and the shape of the diffusion curves. If implementation dynamics are not accounted for, we would be making a big mistake: the difference in the value of research between dynamic and static analysis was £0.84 million for the study on EIS predictive ability and £1.58 million for the study on the progesterone treatment effect.

Obviously, the time at which research reports, in fact the timing of any implementation measure, affects the value of that research study or implementation measure. In this example, the EVRs of the different research studies increased considerably with the dynamic analysis compared with the static analysis, even when the timing of research was considered. A recommendation with research decision would therefore be more valuable when diffusion was considered than with the static analysis. Whether the value of research would cover the cost of these research studies would have to be assessed in further analysis. Ideally, the dynamic EVSIM analysis would be compared with a dynamic cost analysis, accounting for costs when they occur.

7.4.4 Exploring the impact of different study results on the EVSIM

In this thesis, the impact of different study results on the EVSIM and subsequently on the EVR has been largely ignored. The reason for this was to capture the effect of a reduction in uncertainty on diffusion, as was done in previous studies (Willan and Eckermann, 2010). However, a full consideration of the effects of research on diffusion must include the potential for different study results. When this is considered, diffusion is no longer independent of the net benefit function and Equation (4.11) described in Chapter 4 must be used to calculate EVSIM and EVR instead of Equation (4.12).

Implementation of this requires estimates of different diffusion curves conditional on study outcomes. Ideally, a meta model of diffusion as a function of study outcomes would be available. Such a model could be informed by an elicitation of expert opinions on diffusion in which a functional relationship between study outcomes and diffusion is established. To inform this functional relationship it is necessary to elicit a number of different diffusion scenarios conditional on different study outcomes, as was described in Section 6.5.

In this section, I explore the impact of different study results on the Research EVSIM and EVR using an extension to the model shown in Section 7.3.4 and using both proposed studies, on the predictive ability of EIS and the progesterone trial for illustration purposes. For this, I use the individually elicited diffusion curves from each of the three experts as a proxy to diffusion curves for the following three cases:

1. study results are in the lowest quartile of the original distribution, i.e. the predictive ability of EIS is much worse than the previously expected value / the progesterone treatment effect relative to no treatment is worse than for CLS.

2. study results fall within the two mid quartiles of the original distribution, i.e. the predictive ability of EIS falls into 50% credible interval of the previously expected value / the progesterone treatment effect relative to no treatment falls into the 50% credible interval of the previously expected value.

3. study results fall in the uppermost quartile of the original distribution, i.e. the predictive ability of EIS / progesterone treatment effect is much better than the previously expected value.

Implementing Equation (4.11) requires applying the different diffusion curves to each run of the PSA, based on the simulated study results in that PSA run. For illustration purposes, I simplify this by using a different with-research diffusion curve conditional on the expected net benefit associated with EIS. For this, I divide the net benefit distribution up into

quartiles. If net benefit in a PSA run falls into the lowest quartile, I use expert 1's diffusion curve; if it falls into the middle two quartiles, I use expert 2's curve and expert 3's curve for net benefit falling into the uppermost quartile. The counterfactual without research diffusion curve is that based on all (pooled) experts' estimates. The diffusion curves relating to the other two technologies are calculated by using an equal share over the remainder of the market share after EIS has been taken up.

Results of this analysis show that the conditional EVSIMs on study outcomes for both the progesterone and the EIS studies are larger than those calculated independently of the study outcomes (Table 7.11).

Table 7.11. EVSIM of studies with uptake conditional on simulated trial outcomes

Period	Cond. EVSIM Prog trial	Cond. EVSIM EIS trial	Previous EVSIM Prog trial	Previous EVSIM EIS trial
1	£53	£40	£48	£32
2	£156	£129	£123	£82
3	£296	£264	£212	£142
4	£426	£395	£281	£188
5	£504	£471	£301	£200
Average	£287	£259	£193	£129

This in turn leads to the following values for the conditional EVSIMs and EVRs accrued over the population and over time (Table 7.12). The expected value of research accrued for the population and the time horizon is now larger than originally, when study outcomes were not considered in the calculation of the Research EVSIM (£4.07m and £3.23m for EIS and Progesterone studies, respectively). Of course, these results come with the caveat that diffusion curves conditional on study outcomes were not elicited. Furthermore, the net benefit distribution was used as a proxy for study outcomes.

Table 7.12. Accruing conditional EVSIM and EVR over population and time

Study	Per person	Population per annum	Population over decision relevance horizon of 5 years (discounted)
Cond. EVSIM (EIS)	£259	£877,766	£4.39m
Cond. EVSIM (Pg)	£287	£997,430	£4.99m
EVR (EIS)	£311	£1.3m	£6.54m
EVR (Pg)	£288	£860,968	£4.99m

7.5 Discussion

7.5.1 Findings

In this chapter, I illustrated the application of the dynamic expected value of information and implementation method in an exemplary case study on electrical impedance spectroscopy for use in preterm birth screening. This analysis can help assess the losses the health care payer incurs when there is decision uncertainty and predicted low implementation of the cost-effective health technology. Findings of this chapter include that the value of research may be larger than what is commonly assessed with expected value of sample information analysis because the latter typically ignores the effect of research on implementation. When this is accounted for, the value of research can be larger than the EVSI, as was shown in this case study example. When EVR and EVSIM are accrued for the England population, the estimated number of affected patients should be adjusted by proportional uptake. Results of the dynamic analysis differed considerably from the results of the static analysis which means that ignoring implementation dynamics in value of implementation and information analysis leads to erroneous estimates. The timing of research and implementation measures and their effects also alters results and needs to be considered in the dynamic value of information and implementation analysis.

Further findings relate to the case study technology and decision problem. EIS for use in PTB screening together with progesterone treatment of women that test positive was shown to be a dominating strategy against the CL scans and progesterone treatment strategy and against the no screening, no treatment strategy. This was because each PTB averted presented a cost-saving to the payer that, even when few cases were averted, exceeded any expenditure incurred on the purchase of the equipment and on progesterone treatment. Furthermore, EIS appeared to identify more women that would go on to have a PTB than CL

scans, enabling them to receive preventative treatment and thus reduce the number of PTBs. The cost-effectiveness results were largely insensitive to the price of EIS. EIS would be cost-effective up to a capital cost for each EIS device of almost £1 million.

The dynamic EVP was large, at potentially £650 million for the England population over a five year decision relevance horizon. This was caused by low implementation of EIS resulting in a large EVPIM (£631 million). The identified research studies could be worthwhile undertaking if their cost fell below the expected value of research of £3.23 million and £4.07 million for the studies on EIS and progesterone, respectively. When the study results report with a delay in time, these values would be smaller. Compared with the EVP, the EVRs of the two identified research studies only achieved a relatively small reduction in the EVP (of 1-2% of the EVP). The largest part of the EVP was made up of the EVPIM. This points to the fact that there must be other barriers to implementation that have a larger impact than the uncertainty that is contributed by the identified parameters.

7.5.2 Relevance

This chapter illustrates an extension to the EVII framework presented by Fenwick et al. (2008). The present work built on their framework by allowing states of information and implementation other than 'current' and 'perfect', by allowing research to have an effect on implementation and by considering the dynamics and the timing of implementation. This is foreseen to be of value to decision-makers, analysts and manufacturers who wish to perform such analyses. The findings of this case study are relevant to the developer of EIS and decision-makers that may assess EIS for PTB screening in the future. The cost-effectiveness findings for EIS for the use in PTB screening, albeit premature, support the developer's hypothesis that EIS screening together with progesterone treatment could help reduce the incidence of PTBs. As such, this model can be further developed to help the developer build a case for EIS when it is scrutinised by different stakeholders, including potential adopters, once the technology is introduced. It may further be in the interest of the developer and the health care payer to help fund implementation measures or further research to reduce the large burdens of uncertainty and implementation.

The EVRs of the respective research studies were low in comparison to the EVP. There may be different reasons for this: 1. that the main barriers to implementation are not evidence-related; 2. that the study design was not optimal in terms of improving the EVSI and EVSIM such that these studies contributed only little to an increase in implementation; and 3. that there is other evidence that could result in a greater increase in implementation.

The latter would indicate that there potentially is a distinction to be made between the evidence required for demonstrating cost-effectiveness of a new technology and evidence required to promote implementation. An example could be that larger trials or a greater number of trials would be required for promoting implementation than what may be sufficient for demonstrating cost-effectiveness, or vice versa.

Findings also need to be viewed in the context of the evidence that is currently available. Diffusion was elicited under the assumption that the evidence for EIS sensitivity, specificity and progesterone treatment effect would be similar to the evidence available from the pilot study and the evidence available from progesterone after CL scan research studies. If new trials produced different results, this may impact on implementation. A closer examination of other barriers to implementation could also be worthwhile. These have been identified in the qualitative study and include the lack of clinical guidelines, junior staff on rotation, as well as financial issues associated with the purchase of new equipment, amongst others.

7.5.3 Strengths, limitations and future research

The strength of this chapter was the incorporation of diffusion in EVII analysis, the estimation of which was based on the theory of diffusion of innovative technologies (Rogers, 2003) and the Bass model of new product growth (Bass, 1969) via a method of eliciting expert opinions that was proposed in Chapter 6. The extended dynamic EVII analysis was illustrated in a case study of a technology that is still in development. This is relevant because technology assessments commonly occur before technology introduction and when there is still funding for research. I argued that addressing the issue of potentially low implementation at the time of HTA would save the health care system resources and therefore provide health gains for patients. This notion has been quantified in this chapter.

The main limitation is that EVR results do not account for changes in study outcomes. I attempted to illustrate such an analysis in Section 7.4.4. Due to there not being any elicited information on diffusion conditional on study outcomes, these results are only illustrative in nature. There are a number of limitations that relate to the case study model used, especially in terms of simplifying assumptions that were made, all of which have been described in the previous sections. I also did not take side effects of tests and treatments into account. Positive EIS test results may have adverse effects, including anxiety, on mothers, and therefore could cause problems in the pregnancy. This could work in favour of the no screening, no treatment strategy. However, there was no evidence to support this but it should be considered when this model is developed further. In some cases, I had limited data

and based my parameter inputs on assumptions. For further development of the model, a more thorough search of the literature on the predictive ability of CL scans in high risk women and the treatment effect of progesterone in high risk women and after being screened with EIS in particular would be valuable, although the chance of finding data may be slim. Alternatively, elicitation of expert opinion could be performed to obtain better estimates than the ones I have used.

A further limitation is that this chapter focused only on the number of attainable versus desirable purchases, as opposed to the utilisation by clinicians and acceptance by patients. In practice, the value of implementation would need to take all of these into account. Implementation strategies could then be devised to address the low number of purchases and the low utilisation separately. Of course, some implementation strategies would have an effect on both quantities. To simplify the illustration in this case study, I therefore ignored utilisation, but further research should be done on this and on the effect of research studies on utilisation and number of purchases.

The restriction of the economic model to an NHS perspective could also be considered a potential limitation, given that there are likely to be obvious costs to patients, family and society that go beyond the health-care sector. The main argument against using a broader, societal, perspective was that because most assessments of cost-effectiveness are performed from the perspective of the healthcare payer, designing the economic model from the NHS perspective remains most relevant to facilitate comparison with other uses of healthcare resources.

Further research could concentrate on identifying other implementation measures and quantifying their effect on implementation using expert opinion. The extension of the model to all pregnant women rather than women with a history of PTB may be worth doing, given that the findings of this study are encouraging. Of course, such a model would require knowledge of the predictive ability of EIS in that population, something that may be examined in future research studies. Expected value of information analysis on that extension of the model can quantify the value of such research studies.

7.5.4 Conclusion

In conclusion, this chapter showed that diffusion estimates can be included in the EVII analysis framework and can help provide more realistic estimates of the value of research and the dynamic values of information and implementation than available to date. EIS for

use in PTB screening was shown to be a cost-effective, even dominating strategy based on preliminary evidence, but implementation was predicted to be low, resulting in a large burden to the health system. The evaluated research studies alone were not likely to resolve this burden.

CHAPTER 8. THE DYNAMIC COST-EFFECTIVENESS ANALYSIS MODEL ILLUSTRATED IN A HYPOTHETICAL EXAMPLE

8.1 Background

In Chapter 7, I illustrated the dynamic EVII framework in the case study on EIS for preterm birth screening. The aim of the present chapter is to explore the impact of the dynamic cost-effectiveness analysis (DCEA) model that incorporates the experience curve, which was developed in Chapter 4. Originally, this approach was going to be applied in the same case study of EIS for PTB screening. However, Chapter 7 showed that the cost-effectiveness results for EIS were insensitive to price. This results in the experience curve model having no effect on cost-effectiveness results. Based on the current parameterisation of the EIS model, the relationship of implementation and price dynamics with cost-effectiveness cannot be fully explored. Consequently, an alternative parameterisation is used to examine this relationship. With the resulting model no longer representing EIS, cervical length ultrasound, or indeed PTB screening, this chapter presents the application in a hypothetical example in which Technology T1 is compared with Technology T2 and no screening (T0). The flow of the thesis and the place of this chapter in it are shown in Figure 1.2 in Chapter 1.

8.2 Application of the DCEA model in hypothetical example

The hypothetical example used in this chapter is based on the decision tree model for PTB screening presented in Chapter 7. While many parameter values, and hence the evaluated interventions, changed, it was not necessary to change the model structure. The new hypothetical screening technology (T1) used in this example is evaluated against technology T2 and T0 (no screening). The parameter estimates that were changed compared with the model in Chapter 7 are shown in Table 8.1.

Modelling the experience curve requires information on diffusion and on the price change that is associated with diffusion. I used the diffusion parameters obtained from the elicitation study described in Chapter 6 for the ‘natural’ diffusion curve, that is, without

performing further research. The parameters informing the price change element of the experience curve were based on assumptions. For experience curve alpha, I assumed a value of 0.9, such that it fell in the range reported in the study by Brown et al. (2008). This value means that the price of T1 drops to 90% of its previous price every time the initial production quantity doubles. I assumed the initial production quantity at ten produced units, based on the demand predicted by the diffusion curve for the first period plus an additional 50% of stock in case demand exceeded expectations. Another important piece of information for this dynamic analysis was the decision relevance horizon. Despite evidence for very short technology life spans in medical devices (Chapman et al., 2014), I chose a mean of 15 years and varied the decision relevance horizon between 1 and 60 years to explore its effect on cost-effectiveness.

Table 8.1 Parameters used in hypothetical example to illustrate the DCEA model

Parameters	**Mean values**	**Reference**
T1 sensitivity	0.44	Assumption
T1 specificity	0.72	Assumption
T2 sensitivity	0.4	Assumption
T2 specificity	0.7	Assumption
Cost per T1 screening	£ 168	Assumption
Cost per T2 screening	£ 31	Assumption
Cost of care if condition not prevented	£ 3,000	Assumption
Cost of care if condition partially prevented	£ 1,000	Assumption
Cost of care if condition diagnosed and prevented	£ 0	Assumption
Experience curve alpha	0.9	Assumption based on Brown et al. (2008)
Experience curve initial production quantity	10	Assumption
Uptake parameter p	0.018	Elicitation study
Uptake parameter q	0.491	Elicitation study
Uptake parameter m	140	Elicitation study
Decision relevance horizon	15	Assumption

Operationalising experience curve dynamics over the decision relevance horizon required an additional modelling step: the calculation of prices based on the implementation estimate for each future period. In this instance of a fairly short-term decision tree model, each period (in years in this example) also represents one future cohort. In other, more long-term conditions, several patient cohorts could potentially receive the new technology each year. In that case, price changes in each year have to be applied to multiple co-existing cohorts for technologies in which costs occur in more than one cohort. For many medical

devices, which the experience curve model is most relevant for, costs of medical devices often occur at one point in time. In such cases, a period equals a cohort. In situations where costs of the device are incurred for more than one year, a period equals multiple cohorts. Price changes are not only incurred for future cohorts, but also for the first cohort in future periods.

For intervention T1, price changes in each year are a function of implementation in that year. The decision tree then incorporates the price change to calculate the new cost for each year. The costs and effects for each year are recorded up to the decision relevance horizon and are discounted. The average cost (the sum of each year's costs divided by the employed decision relevance horizon) is then the final cost associated with the respective technologies. If there was no discount, the cost per person associated with T2 and T0 would be equal with or without the dynamic model. I assumed that there were no future price changes for technologies T2 and T0. This makes sense for T0 which represents no screening. It also makes sense for the comparator T2, if it is assumed that it is a well-established technology that is predominantly in use for other conditions – even if there were a price change resulting from the additional use for this hypothetical condition, this would likely be negligible.

I first present results from this hypothetical example without the modelling of the experience curve – which I call the static results. Then I present results for the dynamic example with the experience curve; explore the potential effects of diffusion and experience curve parameters on the technology price; illustrate the effect of the decision relevance horizon on the ICER and finally analyse decision uncertainty in the dynamic model.

Both, the static and the dynamic models were programmed in Excel. To represent uncertainty about the parameter estimates, a PSA with 1,000 iterations was performed. The PSA was conducted in Excel using VBA. The uptake parameters p and q were sampled from lookup tables that were generated by sampling from the probability distributions of elicited quantities and by fitting appropriate p and q values. These did not fit any parametric distribution and using the lookup tables was therefore deemed the best way of representing the uncertainty about the mean.

Uncertainty without and with incorporation of the experience curve was analysed using the EVPI. EVPPI analysis was performed to analyse whether uptake and experience curve parameters and the decision relevance horizon were drivers of decision uncertainty in the dynamic model. EVPPI results were obtained using GAM regression and Gaussian Process regression (Strong et al., 2014).

8.3 Results of the DCEA model applied in the hypothetical example

8.3.1 Static results of hypothetical example

Technology T2 was the technology with the highest expected net benefit in this example due to its cheaper price, despite producing slightly fewer QALYs than T1. Table 8.2 presents the costs and effects associated with all the technologies in the static analysis of the hypothetical example. Whilst T0 was cheaper than both T1 and T2, it was also significantly less effective resulting in ICERs that fell below the threshold when comparing both T1 and T2 with it (Table 8.2).

Table 8.2 Cost-effectiveness data in hypothetical example – static analysis

Threshold = £20,000 / QALY	**T1**	**T2**	**T0**
Expected Costs	£1,227	£1,091	£769
Expected QALYs	22.33	22.32	22.27
ICER T1 against…	-	£28,243/QALY	£8,401/QALY
ICER T2 against…	-	-	£6,479/QALY
Expected Net Monetary Benefits	£445,332	£445,371	£444,699

These results are associated with uncertainty. The CEAC in Figure 8.1 shows that from thresholds above £9,000 per QALY there is uncertainty mainly between technologies T2 and T1. This is further illustrated in the cost-effectiveness plane of T1 against T2 in Figure 8.2 in which part of the joint costs and effects distribution is located to the North-West of the threshold value diagonal, and part to the South-East, indicating uncertainty as to whether T2 is truly the most cost-effective technology. The EVPI also indicates decision uncertainty valued at £112 per person which translates into approximately £3 million per annum (assuming a patient population of 27,000 patients, as in the preterm birth model in Chapter 7).

Figure 8.1 CEAC in hypothetical example - static analysis

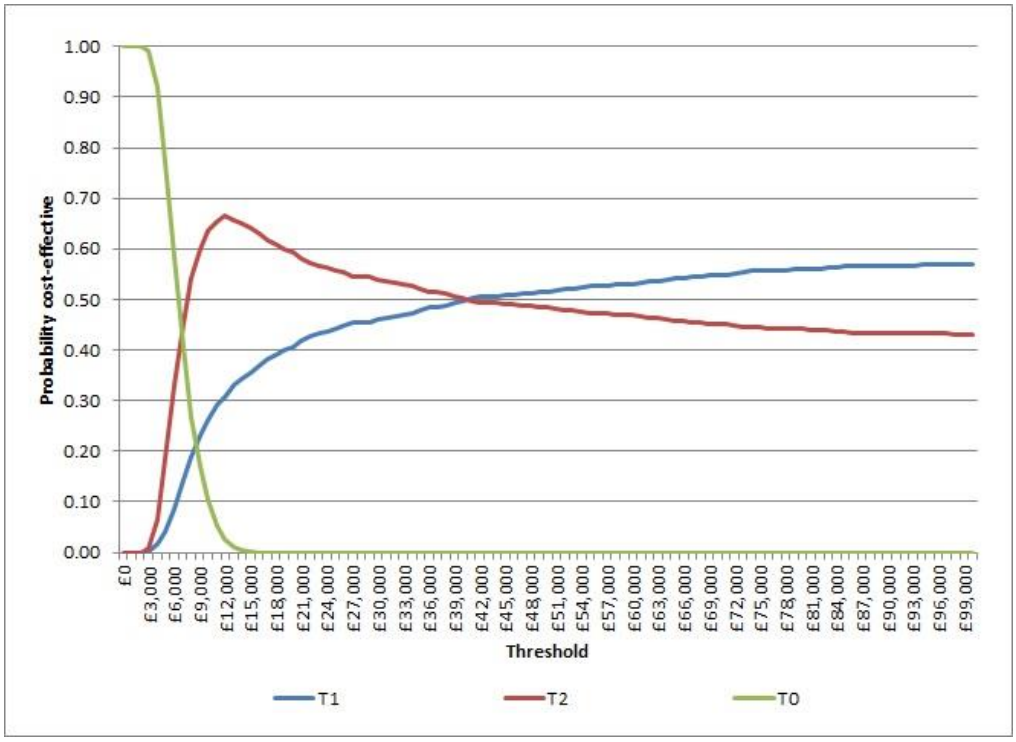
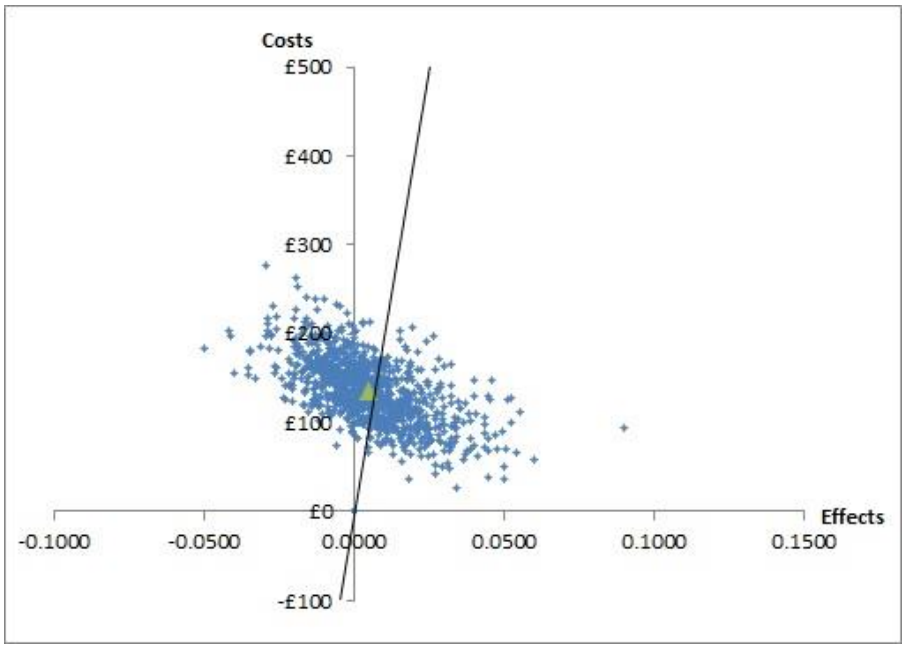


Figure 8.2 Cost-effectiveness plane of T1 vs T2 in hypothetical example - static analysis



8.3.2 Dynamic results of hypothetical example

8.3.2.1 Cost-effectiveness analysis results

In the dynamic analysis, costs and effects have decreased for all three interventions due to the discounting of future periods (Table 8.3). For this reason, effects of the experience curve model on the cost-effectiveness data are not obvious but it can be seen from the expected net monetary benefits in Table 8.3 that the decision between T1 and T2 is closer in the dynamic analysis than in the static analysis.

Table 8.3 Cost-effectiveness data in hypothetical example - dynamic analysis

Threshold = £20,000 / QALY	T1	T2	T0
Expected Costs	£967	£877	£618
Expected QALYs	17.98	17.97	17.93
ICER T1 against…	-	£23,044/QALY	£8,011/QALY
ICER T2 against…	-	-	£6,531/QALY
Expected Net Monetary Benefits	£358,564	£358,576	£358,041

The expected opportunity loss associated with decision uncertainty as quantified by the EVPI is smaller in the dynamic than in the static analysis (EVPI of £103 per person, or £2.8 million for the patient population). This may seem paradoxical but is explained by the discounting of future periods' costs and effects that reduces the expected net monetary benefits of the dynamic compared to the static analysis. The EVPI, being a function of the incremental net benefits and the number of times one would make the 'wrong' decision, consequently decreases. This, however, does not mean that we are more certain about what the truly best decision option is; it just does not matter as much because the cost associated with a wrong decision is now smaller. The CEAC shows that T1 becomes the technology most likely to be cost-effective at a threshold that is lower than that in the static analysis (approximately £28,000 as opposed to £40,000 per QALY) (Figure 8.3). The shape of the joint distribution of the incremental costs and effects of T1 against T2 in the cost-effectiveness plane is altered with the dynamic analysis (Figure 8.4) compared to the static analysis. Incremental QALY gains are now less variable, owing to discounting of future periods. The wider spread of incremental costs is caused by the inclusion of the experience curve which in some of the iterations of the PSA results in T1 being cheaper than T2 (Figure 8.4).

Figure 8.3 CEAC in hypothetical example - dynamic analysis

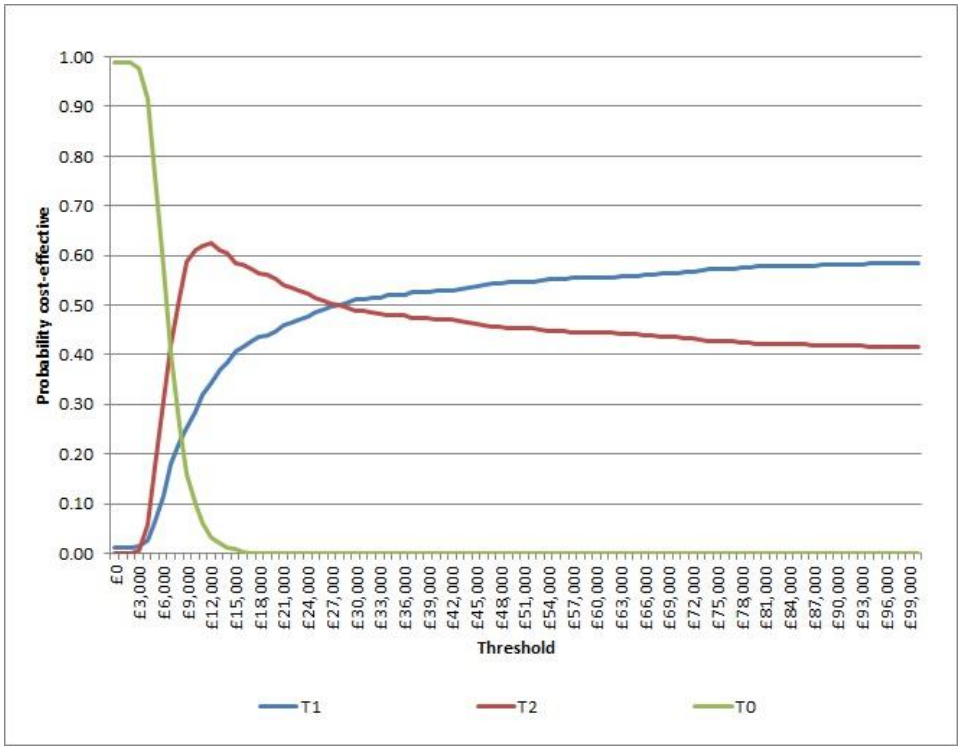
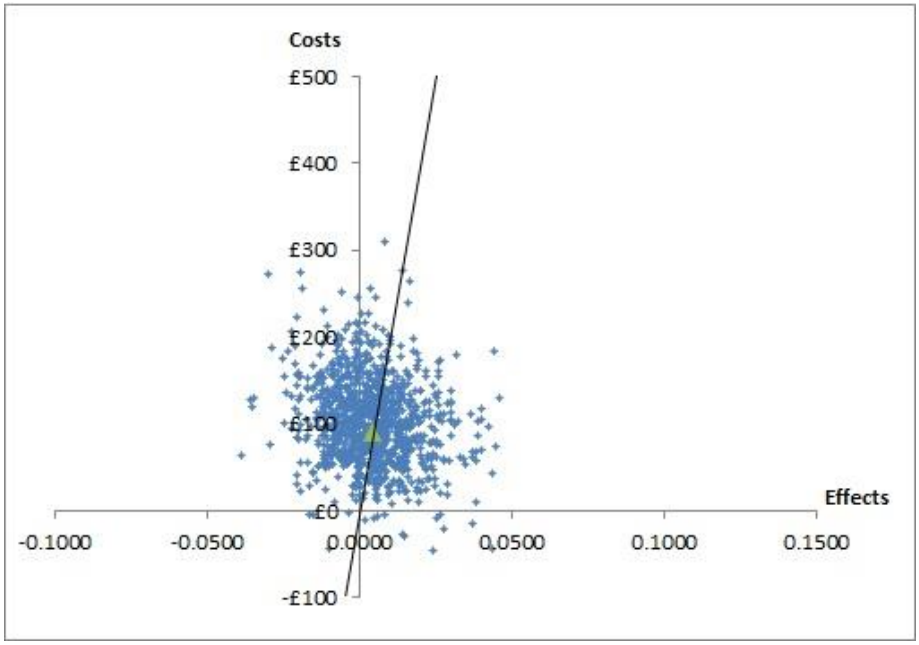


Figure 8.4 Cost-effectiveness plane of T1 vs T2 in hypothetical example - dynamic analysis



8.3.2.2 Exploring the potential for future price changes in the hypothetical example

This hypothetical example shows the potential shape of price declines when future periods with price changes contingent on implementation dynamics are modelled (Figure 8.5). Figure 8.5 shows two different developments: one is that cumulative diffusion of the new screening technology T1 follows an s-shaped curve, as was already shown before. The maximum number of attainable adoptions is reached after approximately 15 years, emphasising the word 'attainable' as opposed to 'desirable'. The second development is that price starts to decline after approximately 12% of the attainable adoptions have been achieved (Figure 8.5). Price remains stable for a short period of time (two years), which is followed by a quick price decline. With uptake exhibiting diminishing marginal growth towards the later periods, price converges to an asymptote of £3,348 after 22 years. More intuitively, when uptake growth slows down, the reduction in technology price decreases until the lowest possible level of price is reached. Price is shown as a function of the cumulative number of adoptions in the traditional experience curve chart in Figure 8.6 (Brown et al., 2008). These developments are explained by the parameterisation in this example, particularly the initial production run, and the experience curve alpha and diffusion parameters. The initial production run was assumed to produce small quantities (10 devices in the first year) in response to the predicted low uptake. With uptake growing quickly in this example, the initial production run is used up in the second year; that is, at the time at which implementation levels increase exponentially.

Figure 8.5 Price and uptake developments in hypothetical example

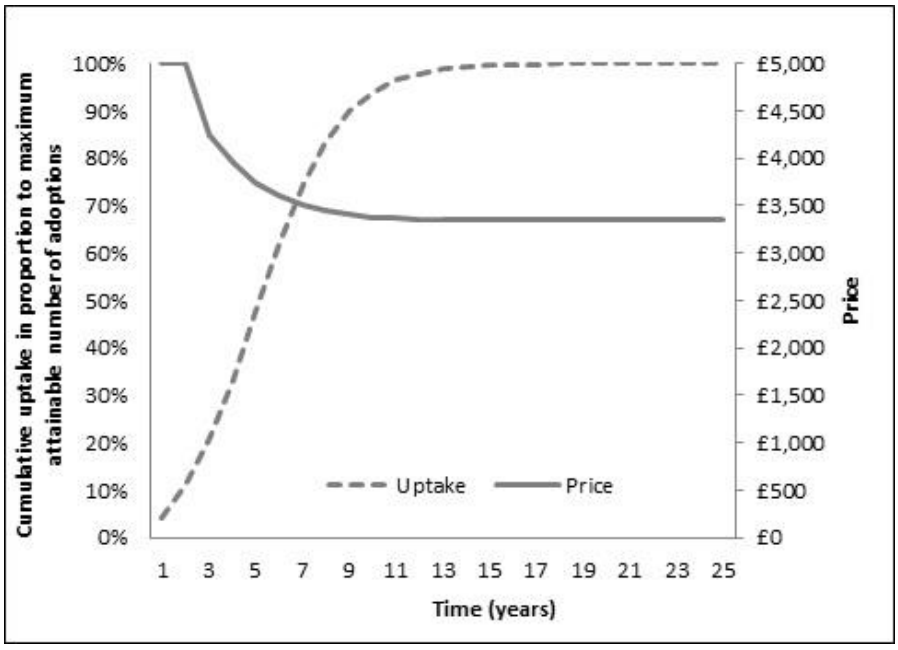
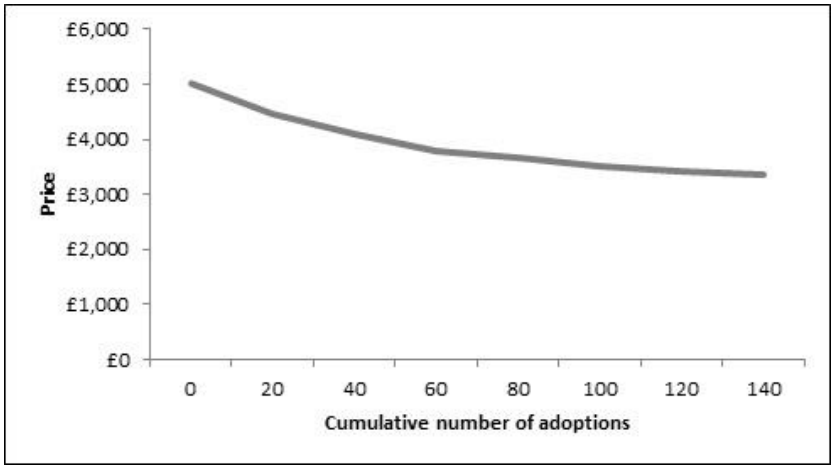


Figure 8.6 Traditional experience curve chart for hypothetical example



With different parameter values, such as larger initial production runs or slower diffusion curves, the price decline could be less drastic. I show examples with a larger initial production run of 65 produced units and the same diffusion curve in Figure 8.7 a), and small and large initial production runs (10 and 65 units) with a much slower diffusion curve with parameters taken from ultrasound scans (Gobok et al., 2009) in Figure 8.7 b) and c), respectively. These graphs show that price declines could happen later in the lifetime of a new technology and may be less pronounced with different initial quantities produced.

Figure 8.7 Potential impact of parameter values on price and uptake: initial production run and diffusion

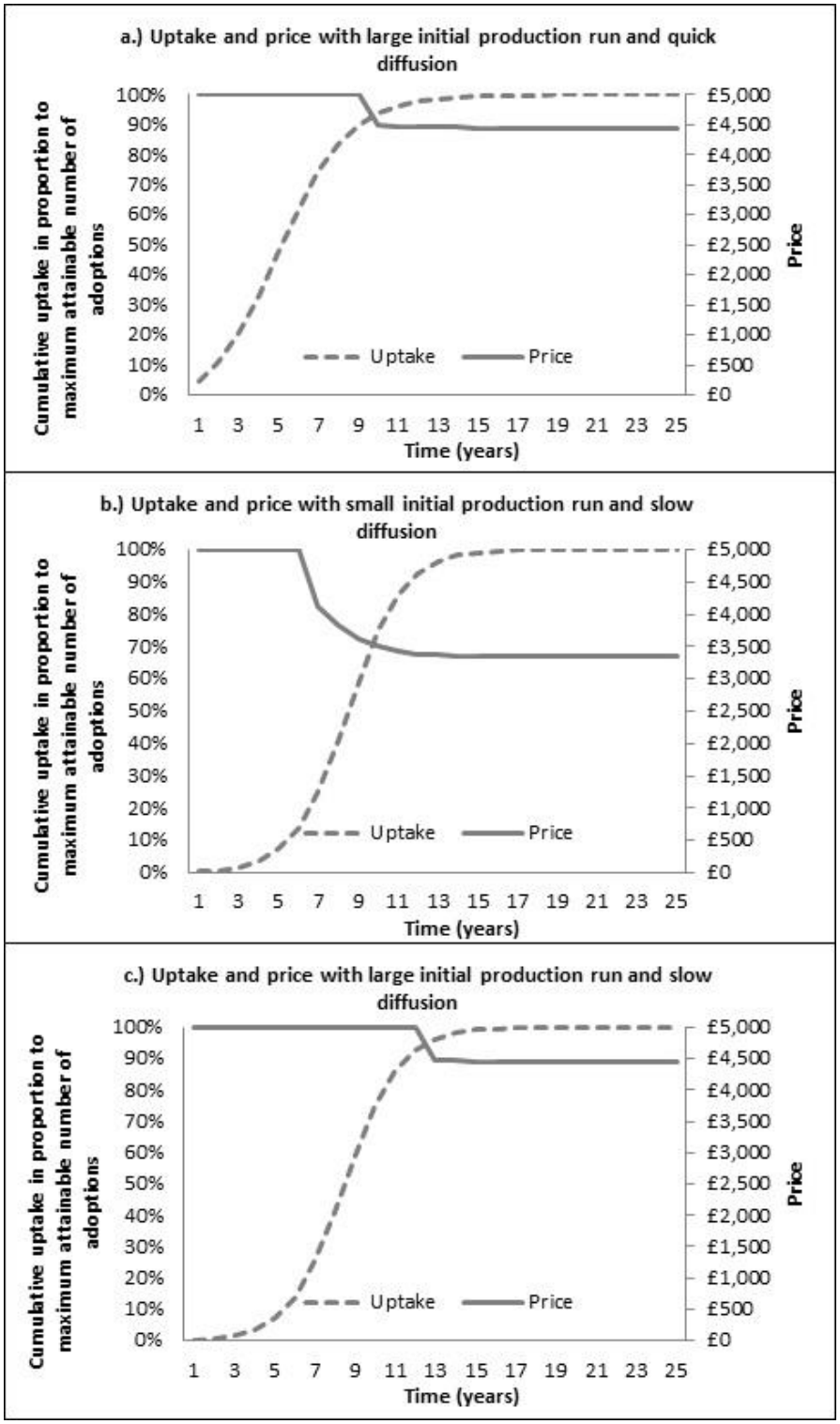
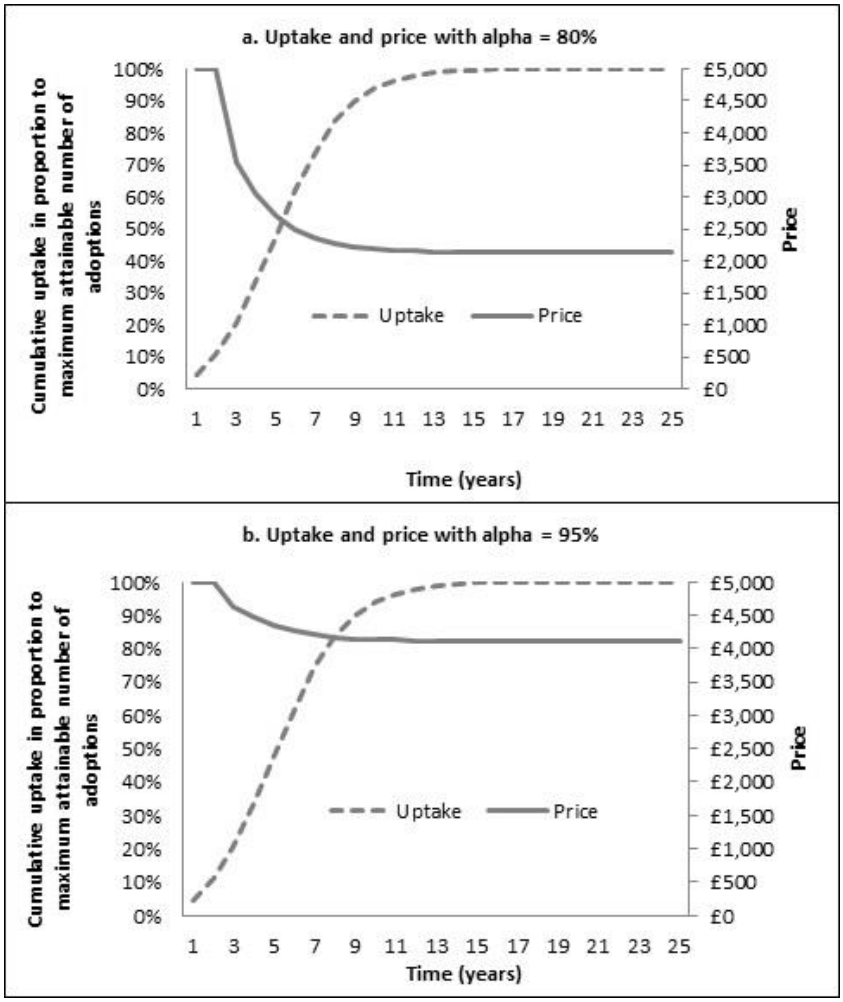
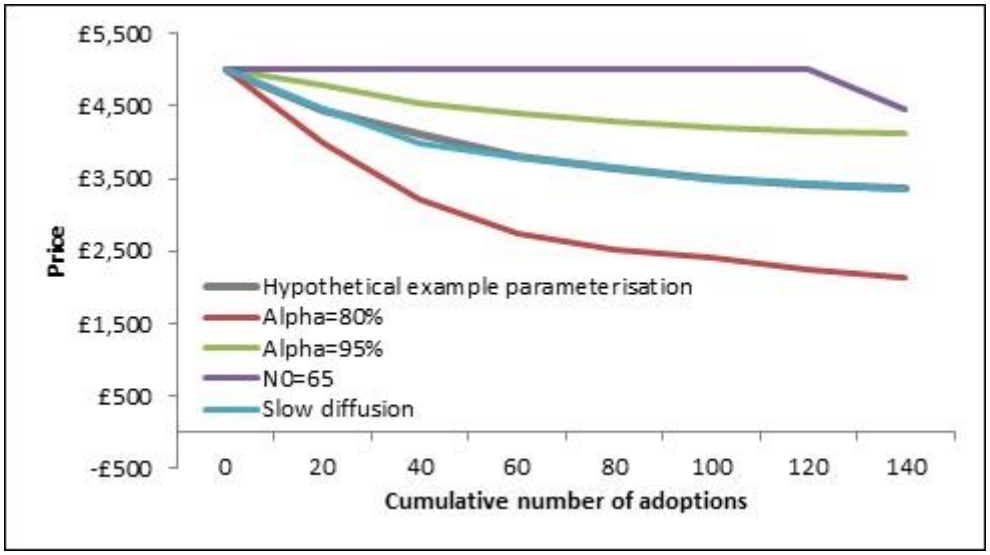


Figure 8.8 Potential impact of parameter values on price and uptake: experience curve alpha



Choosing different magnitudes for the experience curve parameter alpha, which is the percentage to which price declines every time initial production quantity doubles, can result in large differences in the magnitude of price reduction that can be achieved. This is illustrated in Figure 8.8, where, given that all else remains equal, an alpha of 80% could reduce future price to less than half of its starting value while an alpha of 95% would still reduce the future price to just more than 80% of its starting value (Figure 8.8) (noting that the initial production run was very small in this hypothetical example). The effects of these different variations to the parameter values are also shown in Figure 8.9 in the traditional experience curve chart, in which price is plotted against number of adoptions instead of time. Of course, slow diffusion makes no difference to the experience curve shape, which is the rationale for a presentation in a chart that shows both, price and uptake developments as, for instance, in Figure 8.5.

Figure 8.9 Traditional experience curve chart for variations in parameterisation



The significant effect that the experience curve alpha parameter has on the experience curve also translates into different ICERs. This is shown in Table 8.4, where I present the resulting ICERs when different sizes for experience curve alpha are assumed. Alphas between 70% and 95% could make a difference of almost £4,000 to the ICER in this example.

Table 8.4 ICERs with different values for experience curve alpha

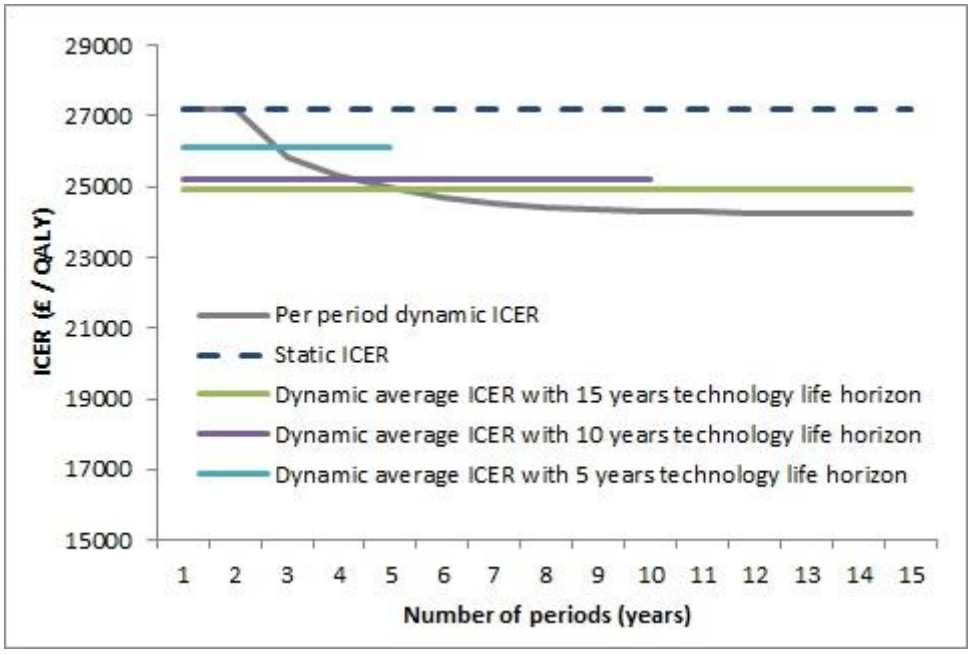
Experience curve parameter alpha	Deterministic ICER of T1 vs T2 (per QALY gained)
60%	£21,310
70%	£22,204
80%	£23,433
90%	£25,068
95%	£26,063
100% (no change in price)	£27,190

8.3.2.3 The effect of the decision relevance horizon on the dynamic ICER

As was shown above, the extent to which price changes occur largely depends on the parameter values of the pivotal diffusion and experience curves. These effects translate into changes in cost-effectiveness. In addition to those effects, the dynamic ICER is affected by the decision relevance horizon which crucially determines how much lower the dynamic ICER is compared to the static ICER. Figure 8.10 shows different dynamic ICERs (which are the averages over the respective decision relevance horizon) plotted against the number of periods up to the chosen decision relevance horizon. A decision relevance horizon of 15 years results in the lowest dynamic ICER compared to decision relevance horizons of five and ten years. This negative relationship between the dynamic ICER and decision relevance horizon exhibits diminishing marginal returns. This is best explained by the per-period dynamic ICER decreasing with diminishing marginal returns (Figure 8.10). The diminishing rate is caused by prices stagnating at a certain time period, with the effect of the per-period ICER not being significantly lowered by modelling more periods in the future.

When a decision relevance horizon of 15 years is chosen, the dynamic ICER is lower than the static ICER, at just below £25,000 per QALY as opposed to the static ICER of just above £27,000 per QALY. Choosing a shorter decision relevance horizon of five years, in contrast, may mean that price changes have not yet been realised and that the dynamic ICER remains closer to the commonly used static ICER at just above £26,000 per QALY (Figure 8.10). While the decision relevance horizon is not a factor that alters the decision in this hypothetical example (if the threshold is £20,000 per QALY), this analysis shows that the employed decision relevance horizon could be influential when the experience curve is modelled in cost-effectiveness analysis.

Figure 8.10 ICER results in hypothetical example with and without experience curve



8.3.2.4 How does the addition of experience curves affect decision uncertainty in the hypothetical example?

As was mentioned above, the EVPI reduced with the dynamic analysis compared with the static analysis (£103 and £112 per person, respectively). This was explained by discounting of future periods. Whilst it makes sense to value future price less than the starting price in the first period, I explored the expected opportunity loss associated with decision uncertainty when future periods were not discounted for illustration purposes. The resulting EVPI was £124, which is larger than the EVPI resulting from the static analysis.

The analysis including discounting of future periods resulted in diffusion parameters accounting for approximately 4% of the EVPI at £6 per person (Table 8.5). The EVPPI of the experience curve parameter alpha is only worth £0.01, and that of the decision relevance horizon £0.06. Together, the diffusion and experience curve parameters have a grouped EVPPI of £9 and the diffusion and decision relevance horizon parameters a grouped EVPPI of £10. Diffusion parameters *p* and *m* have an individual EVPPI of £3.55 and £3.64, respectively; diffusion parameter *q* only has an EVPPI of £0.64. Standard errors of the EVPPI are relatively large, indicating difficulties in estimating the EVPPI that may be caused by the nonlinear nature of this model. The large standard errors may explain why the grouped EVPPI of all the newly introduced parameters is smaller than the grouped EVPPI of a subset of four of them. Furthermore, 1,000 simulations of the PSA may not be sufficient to provide accurate estimates but model run times were prohibitively long. Lastly, the

estimation method used was different for the five parameters: instead of the GAM regression, the Gaussian process regression was used because the GAM regression becomes increasingly inaccurate as the number of parameters increases (Strong et al., 2014).

Table 8.5 EVPPI results for hypothetical example - dynamic analysis

Parameter(s)	Per person EVPPI (£)	Standard error	Percentage of EVPI (%)
Grouped diffusion parameters	6	3.8	4
Experience curve parameter α	0.01	1.62	0
Decision relevance horizon	0.06	1.79	0
Grouped diffusion and experience curve α parameters	9.4	4.05	9
Grouped diffusion and decision relevance horizon parameters	10.4	4.37	10
Grouped experience curve α and decision relevance horizon parameters	0.1	2.2	0
Grouped diffusion, experience curve α and decision relevance horizon parameters*	4.13	3.66	4

* For this analysis, a Gaussian process regression was used instead of a GAM regression.

8.4 Discussion

8.4.1 Findings

In this chapter, I have applied the DCEA model in a hypothetical example. This analysis shows that changes in price precipitated by changes in uptake significantly affect cost-effectiveness results and have the potential to alter decisions. The adopted decision relevance horizon plays a significant role in determining whether price changes reverse the cost-effectiveness outcome, as short decision relevance horizons may not be sufficient to allow price changes to take effect. EVPI results were also affected, with the new parameters contributing to decision uncertainty. Furthermore, taking the average of future discounted periods resulted in a lowered EVPI compared with the static analysis. Results of the partial EVPI analysis imply that, at least in this example, there is value in reducing uncertainty surrounding diffusion, experience curve and decision relevance horizon parameters;

although with the caveat that estimating the true EVPPI is made difficult by the nonlinear structure of the model.

8.4.2 Relevance

These results call into question the commonly made assumption of the first cohort being representative of future periods until re-appraisal is undertaken, and the common disregard for changes that are precipitated by the reimbursement decision itself. The proposed model is especially useful in technologies that may be rejected at the common cost-effectiveness threshold but that may exhibit a decline in price with increasing uptake. As such, it may be useful to present this analysis as a scenario analysis in a submission. In technologies, for which price does not represent a substantial part of its cost to the health care system, this analysis may not affect model outcomes considerably. It is necessary to reiterate that I am only aware of empirical evidence supporting the experience curve in medical devices. Whilst such price changes may exist for other technologies, the lack of evidence may mean that the pragmatic decision maker would prefer confining this additional analysis to medical devices. In principle, of course, this model can be used in all technologies, with parameter distributions reflecting whether a price change is expected or not.

It is important to recognise that, when modelling future cohorts, there is a trade-off of present against future welfare. I assumed here that discounted future welfare gains of one technology could offset present welfare gains of another technology. The key problem with this is the uncertainty surrounding future events. Price changes might never materialise or another more cost-effective technology could become available. Careful consideration of competitor technologies to be launched subsequently is therefore advisable. It is furthermore important to treat the experience curve alpha parameter just as any other uncertain parameter and reflect that uncertainty.

8.4.3 Strengths, limitations and further research

The strength of this chapter lies in the exploration of the effects of the experience curve, diffusion and decision relevance horizon parameters. The choice of a decision relevance horizon was shown crucial for the value of the dynamic ICER. There is differing literature on the appropriate time horizon. Hoyle (2010) estimated the mean drug lifetime to be 57 years (95% confidence interval 39-79 years) and used this as a proxy to a time horizon. In

contrast, medical devices seem to have much shorter lifespans, estimated as short as 18 months (Chapman et al., 2014). While the ISPOR Good Research Practices task force (Caro et al., 2012) recommends a time horizon long enough to capture all relevant outcomes, noting that this may result in a lifetime horizon, the interpretation of this refers to the within-cohort time horizon rather than to the number of periods that should be modelled in the future for cohorts that start at different times. No matter what decision relevance horizon is chosen, it is appropriate to include this within the PSA because of the uncertainty associated with its estimates.

A limitation of this chapter is that the DCEA model was illustrated in a hypothetical example only, rather than in a real world example. Furthermore, the model structure was taken from the PTB model presented in Chapter 7 and, despite changes in the parameter values, resulted in technology price that made up only a fraction of the overall cost associated with the screening strategy T1 (6%). This chapter has therefore highlighted that the experience curve approach may be most important in higher cost technologies, and certainly in those in which technology cost makes up for a major part of the overall cost associated with a strategy.

Furthermore, I did not illustrate the EVR here, because this was done in Chapter 7 and the workings of calculating the EVSIM and EVSI are not different in this analysis, with the exception that net benefit is time-dependent. So, for each future period that the EVSIM is calculated for, the net monetary benefit of that period is based on that period's price, which is conditional on diffusion.

For the purpose of illustration, experience curve parameter alpha (that is the percentage of price to which price changes with diffusion) has been assumed to be at 90%. It is not clear what the true value of this alpha would be for EIS, or any technology, apart from those where the actual price change has been empirically estimated. (Brown et al., 2008) reported that alphas in most published experience curves fall into the 70-80% range. Assuming an alpha of 90% was therefore conservative, resulting in a smaller change in the ICER than lower values for ICER, compared to no future price changes. However, the chosen value would crucially affect the shape of the experience curve and consequently the ICER. I addressed the uncertainty associated with it by varying alpha in the PSA. If such analysis were used for decision-making, however, a better estimate of this parameter would be desirable. This could be obtained through elicitation of expert opinion, but ideally a commitment to a price change conditional on uptake would be obtained from the company, in the form of a price-volume agreement scheme. Such a scheme would eliminate substantial

uncertainty for the decision-making body and the NHS, whilst also ensuring that the company only has to lower their price when a certain sales volume is reached.

Further research should be conducted to generate evidence for or against the existence of the experience curve in a greater sample of medical devices; and in other health technologies as well.

8.4.4 Conclusion

In conclusion, the DCEA framework that includes modelling the experience curve has the potential to alter decision-making. The framework is especially relevant for those technologies for which there is empirical evidence of experience curves (that is, medical devices) and those technologies for which price is a main contributor of overall costs associated with the strategy. Furthermore, this analysis is most relevant when the technology in question is not cost-effective and when there is decision uncertainty. Otherwise, future price reductions are not likely to have an impact on cost-effectiveness outcomes.

CHAPTER 9. DISCUSSION AND CONCLUSION

9.1 Summary of findings

In this thesis, I explored the impact that estimates of health technology diffusion have on cost-effectiveness and value of information and implementation analyses. Four hypotheses motivated this thesis: (A) consideration of implementation could be beneficial to the reimbursement authority and the health-care payer at the time of HTA. (B) Implementation estimates may be correlated with other parameters in the cost-effectiveness model and influence health economic outcomes. (C) It is possible to acquire estimates of implementation that are compatible with diffusion theory; and incorporate these in an HTA. Lastly, hypothesis (D) was that a better understanding of the technology-specific causes of low implementation is required to improve implementation.

I tested hypothesis (A) by extending the EVII framework to include dynamic implementation estimates and accounting for the effect of research evidence on implementation. I found that knowledge of low implementation and its associated opportunity loss at the time of HTA presents an opportunity to recommend research studies that serve both the reduction in uncertainty and improving implementation. The opportunity loss incurred by the payer when the most cost-effective technology experienced low implementation and its cost-ineffective comparators remained in use could be assessed using the value of implementation. Furthermore, I found that the value of research is under-estimated with traditional EVSI methods, if the research also has an effect on implementation. The expected value of research that consists of both its value of information and implementation allows for these two separate effects. The value of research may, however, be over-estimated if its calculation involves scaling up over the entire eligible patient population rather than over the population that will receive the technology. If different study results are considered, EVR results may differ further.

Price can be influenced by diffusion and I found that this relationship could be represented by experience curves. Incorporating experience curves in the dynamic cost-effectiveness analysis framework led to potential differences between traditional cost-effectiveness and dynamic results (hypothesis (B)) (see Chapter 8). These differences in cost-effectiveness are caused by price decreasing with increasing uptake, a relationship that has been observed in medical devices.

I demonstrated that it was possible to obtain estimates of future diffusion (hypothesis (C)). Literature reviews presented in Chapter 3 highlighted that methods were needed to estimate diffusion without prior data available. I proposed a method that, with elicitation of three quantities only, can generate diffusion curves over the life-span of a technology. This method was based on diffusion theory and an existing diffusion model. Diffusion curves were elicited for EIS, assuming availability of different sets of evidence.

Effective implementation strategies are best devised if there is knowledge of the technology- and lifecycle-specific factors influencing diffusion, but it can be challenging to identify those strategies with the largest effect sizes without an elicitation of expert opinion (hypothesis (D)). Barriers and facilitators to diffusion were shown to be highly specific to individual technologies. To identify EIS specific diffusion determinants and determine the design of potential implementation strategies, a qualitative study was therefore more appropriate than drawing on an existing catalogue of diffusion determinants (Greenhalgh et al., 2004) or on other studies examining diffusion patterns and determinants (Barnett et al., 2011, Department of Health, 2011).

With respect to the case study on EIS for PTB screening, EIS was shown to be dominating against the alternative cervical length ultrasound scans and against no screening with relatively high certainty, based on current evidence. This outcome was shown to be insensitive to the price of EIS, and the experience curve model was therefore illustrated in a hypothetical example. Relevant implementation strategies to EIS would be related to establishing evidence in the first instance. Diffusion of EIS was estimated for counterfactual with or without research scenarios and it was shown that in all scenarios, EIS would not reach the maximum desirable adoptions, and diffuse over a period of 10 to 14 years. Low implementation resulted in a large expected value of implementation; and the expected value of information was relatively smaller. The expected value of research associated with the two defined research studies may warrant performing these studies, if their value falls below costs. Further implementation measures may need to be considered in order to address low implementation.

9.2 Novel contributions to knowledge

This thesis made four main novel contributions to knowledge:

1. Incorporating estimates of health technology diffusion in health economic methods

This thesis has, for the first time, incorporated a formal technology diffusion model in EVII analysis and cost-effectiveness analysis. In cost-effectiveness analysis, diffusion was ignored and static estimates of implementation were used in the few studies that considered uptake. When this thesis began, methods of obtaining implementation estimates for use in expected value of information and implementation research were rudimentary and often did not account for any dynamics. In fact, in value of implementation analyses, all implementation estimates found through reviews reported in Chapter 2 were of static nature. The counterfactual, that is 'natural' diffusion that would occur without any implementation measures, was also largely ignored. Since then, other studies have been published that included developments of implementation over time (Whyte et al., 2016, Faria et al., 2014, Andronis and Barton, 2016), but none of them used elicited data that was based on diffusion theory. Whilst implementation changes over time were used in EVIM analyses prior to finishing this thesis, this work remains the only one that based diffusion estimates upon theory and established models. The emergence of this other literature highlights the importance and timely nature of this thesis.

2. <u>The expected value of research as a combination of value of information and implementation</u>

I developed the expected value of research that explicitly considers effects of information on implementation. This has not been done before based on the reviews reported in Chapter 2. Within this, this research has also contributed to clarifying a misunderstanding of the realisable EVPI and the implementation-adjusted EVSI: these are measures of the value of implementation and not value of information measures as was suggested in the study by Andronis and Barton (2016). This thesis makes the distinction between value of implementation measures and value of information measures much clearer. I also highlight that the value of research can be larger when both, expected value of information and implementation, are considered than in traditional EVSI analyses.

3. <u>New method to elicit diffusion prior to technology introduction</u>

The method of predicting health technology diffusion developed in this thesis is the first method that does not require data of initial periods or data on analogous technologies to be available. The literature review in Chapter 3 showed that the only methods of forecasting using an established quantitative model without any diffusion data available were reliant on diffusion data of analogous technologies (Goodwin et al., 2014). The use of analogous

technologies is, however, associated with many limitations. First, it is not obvious what discerns an analogous technology and multiple factors need to be assessed in order to decide whether another technology may be an appropriate analogy. Second, the probability of an analogous product being available in the health care industry may be smaller than with many other goods and especially consumer goods, because of patenting that prevents very similar technologies from entering a market. Being able to predict diffusion independently of the availability of diffusion data for analogous technologies is therefore important. The structuring of the diffusion model using the maximum attainable number of adoptions, the number of adoptions in the first year and the time to the inflection point of the diffusion curve will be particularly useful to analysts who wish to model diffusion as an s-shaped curve, with theoretical and empirical foundation.

4. Extended cost-effectiveness analysis framework accounting for changes in implementation and price

While evidence on the existence of experience curves in health technologies was presented before (Brown et al., 2008, Brown et al., 2007), the relevance of these developments to the economic evaluation of health technologies was and remains unexplored by other studies, as was shown in Chapter 2. This work is novel in that it provides the DCEA framework, which enables calculation of ICERs with future periods that account for changes in implementation and price that are precipitated by the reimbursement decision.

9.3 Relevance and implications for decision-making and the manufacturer

9.3.1 Preventing the burden of low implementation

As was demonstrated in this thesis, there could be considerable value in preventing the burden of low implementation to the health system. This could be achieved by assessing this burden and evaluating measures to reduce it at the time of technology appraisal. The benefits that can be made by improving implementation can, in some circumstances, exceed those that can be made by reduction of uncertainty. Assessing this early on may therefore be in the interest of the payer and the decision maker.

Research projects may become more attractive when their value in improving implementation is considered. Whilst reimbursement authorities such as NICE typically do not have the mandate to recommend implementation initiatives, they can recommend further research being conducted if there is large decision uncertainty (NICE, 2013). In practical terms this could have a substantial effect on recommendation with research (RwR) decisions, which are a type of managed entry agreement used by reimbursement authorities and manufacturers to agree on a process of recommending a new technology (Walker et al., 2012a, Walker et al., 2012b, Grimm et al., 2016). With this analysis, the value of research is not limited to its impact on decision uncertainty, which can be estimated by the EVSI, but includes the associated value of increased implementation and may thus be of particular relevance to reimbursement authorities. Of course, a RwR decision requires other preceding analyses (Walker et al., 2012a, Walker et al., 2012b) to ensure it produces a positive net monetary benefit; otherwise the faster implementation can inflate opportunity costs associated with ineffective schemes.

The dynamic EVII analysis presented here can inform the design of evidence generation schemes, thus producing a more appropriate estimate of the value of research to the payer. For instance, it might not be the research scheme with the largest EVSI that provides the highest value of implementation. This can occur when the main drivers of decision uncertainty do not coincide with the main evidence-related drivers of diffusion. Evidence for such a situation was presented in the case study of this thesis where experts interviewed in the qualitative study stated that they perceived the clinical uncertainty surrounding EIS for use in PTB as being large and a perceived barrier to implementation (Chapter 5). The health economic model, in contrast, showed that decision uncertainty associated with EIS for use in PTB was small and related to uncertainty surrounding a combination of many different model parameters instead of mainly the predictive ability of EIS and the treatment effect of progesterone. In such a situation, where there is potential value in different research studies, calculation of the EVSIMs and EVSIs of these research schemes needs to be undertaken and the research project chosen that exhibits the largest EVR relative to its cost. Where the main drivers of uncertainty are the same as the main evidence-related drivers of diffusion, the EVR analysis could still affect the research scheme design through different sample sizes being indicated. To design eligible research schemes for assessment via EVR methods, it is worth watching out for study results of a currently ongoing study on how evidence influences decisions in the UK NHS, including what defines ‘strength’ and ‘credibility’ of research in the eyes of different stakeholders involved in decision-making in the NHS (Turner et al., 2016).

Lastly, performing EVII analysis may lead to the more realistic assessment of evidence generation schemes because the uptake-adjusted population values result in smaller values of research. The difference between the full population values of research and the uptake-adjusted values of research are largest in the first few periods modelled, with the per-period estimates in the later periods converging because of discounting. The uptake-adjustment is hence most relevant when shorter decision relevance horizons are used.

9.3.2 Recommending technologies based on future prices

Analyses that incorporate price changes precipitated by the reimbursement decision and subsequent diffusion may pose a useful scenario analysis for a decision-maker. The proposed DCEA model is especially useful in technologies that may be rejected at the cost-effectiveness threshold but that may exhibit a decline in price with increasing uptake. It can help avoid a situation in which a potentially cost-effective technology is rejected. Such analysis is relevant when price changes are triggered by a reimbursement decision in the jurisdiction of interest and when there is no re-appraisal scheduled for the time when those price changes occur.

The proposed DCEA framework entails a trade-off of present against future welfare and has implications for the ongoing debate on pharmaceutical pricing. Whilst there is support for higher prices for new drugs that reflect the benefit of future innovation (Jena and Philipson, 2007), another view is that such future benefits are uncertain and that gains appear once future prices reduce because of generic competition (Claxton et al., 2008). This framework suggests that these price changes can occur prior to generic competition and as such can support higher prices at product launch. However, the size of the premium at launch suggested by our framework would depend on a number of factors and would require further research. Reimbursement bodies may want to consider reducing the uncertainty around initial pricing and future price changes by making reimbursement contingent on the establishment of price and volume agreements. Whilst the precise mechanism by which this could be enacted is beyond the scope of this thesis, this analysis can help in assessing whether price volume agreements are mutually beneficial by making the relationship between implementation and price over time explicit and enabling uncertainty to be captured.

It is noteworthy that health technology assessments of medical devices and other technologies including diagnostics follow different rules from typical NICE technology appraisals. For instance, the NICE medical technologies evaluation programme requires for

recommended devices to be cost-saving (NICE, 2011b). But these different requirements do not make the proposed analysis redundant: future price reductions precipitated by the reimbursement decision could make a technology that is not cost-saving at current price cost-saving at future price developments and as such the proposed analysis would be relevant to such a setting.

9.3.3 Using diffusion estimates in health technology assessments: considerations for decision makers

To offer pragmatic guidance as to when these proposed approaches are most valuable in the context of HTA, the following conditions need to be met: they are broadly summarised as (1) technologies that experience low implementation (in at least one period) and (2) technologies that experience price changes that are precipitated by and relevant to the recommendation decision. Only the first condition has to be met for expected value of implementation analysis to become relevant and for making population estimates of value of information analysis more realistic. In addition to the first condition, the second condition needs to be fulfilled to warrant calculation of the future periods ICER.

Assessing whether these conditions hold for a new health technology is not trivial. There may be some idea of whether a technology may be prone to low implementation. There may be certain health technologies that are generally less prone to low implementation. To my knowledge, however, there is no standard way of identifying technologies with potentially low implementation, which suggests the need for a screening process that should precede any further analysis and that will help identify future implementation levels. This screening process could entail qualitative interviews with relevant stakeholders and ideally an elicitation exercise, the kind that were described in the previous chapters. Evidence from the diffusion curves review in Chapter 3 showed, however, that none of the health technologies had full instant implementation. Even after NICE guidance with recommendation decisions, no sudden jump to full implementation occurred. For lack of other experience, and in order to avoid the complexities and time and resource requirements associated with preceding qualitative and elicitation studies, I therefore recommend assuming that condition (1) holds for all health technologies.

In the context of value of implementation analysis, there is still a judgement to be made on whether the extent to which low implementation occurs for a technology of interest warrants performing further analyses and potentially conducting qualitative and elicitation studies. Whether the burden on the health system that results from low implementation is

large depends on three factors: a) how low implementation is going to be, that is the implementation discrepancy, b) the magnitude of the incremental net benefit of the new technology over its comparators, c) the size of the population affected. Factors b) and c) will be known to decision-makers at the time of the appraisal. Factor a) could be addressed by using diffusion estimates that are estimated in the context of budget impact analysis. The ISPOR guide to budget impact analysis already recommends that estimates of natural diffusion be submitted (Mauskopf et al., 2007). A full set of value of information and implementation estimates should be produced using these diffusion estimates. This would then allow future research to be prioritised and potentially built into a RwR scheme.

Assessing whether condition (2) holds, that is, whether a technology will experience price reductions that are precipitated by, and are relevant to, the reimbursement decision, is no less complex. Such price reductions may be most applicable to medical devices as there is no evidence on experience curves occurring in pharmaceuticals, although there is evidence on price changes in pharmaceuticals (Hoyle, 2010). Existing evidence on pharmaceutical price changes does not explain whether these would occur independently of the reimbursement decision in the same jurisdiction (Hoyle, 2011). More research on the nature of price reductions in pharmaceuticals may therefore aid in this decision. With experience curves evidence existing only for medical devices, these technologies should potentially be singled out for this analysis until further evidence emerges. Furthermore, the experience curve analysis is only relevant to assessments in which the health economic output, commonly the ICER, is sensitive to technology price. Only then would the DCEA and experience curve analysis influence results, and hence, become relevant for decision-making.

9.3.4 The manufacturer's perspective

This study could contribute to the perspective of the pharmaceutical and medical device industries by providing estimates of health technology diffusion that could inform profit calculations and the return on investment in a clinical trial. The manufacturer's objective function has been described as a profitability index that is a ratio of the net present value of all costs and benefits over the investment costs, amongst other objectives by Phillips and Bana e Costa (2007) in their work on corporate and not-for-profit decision-making. Considering all development options, the manufacturer thus faces a portfolio analysis and option value decision problem whereby the investment in product and evidence development programmes associated with each option have to be accounted for and weighed against

future profits. To be able to forecast future profits for each option, it is essential to know the expected sales volume, which should be informed by diffusion estimates.

Funding for clinical trials comes mostly out of the pharmaceutical industry (Breeze and Brennan, 2014) and several studies have incorporated the pharmaceutical perspective in, commonly health care provider perspective oriented, value of information analyses (Breeze and Brennan, 2014, Willan and Eckermann, 2010, Pezeshk and Gittins, 2006). The study by Breeze and Brennan (2014) is a useful contribution in that the authors calculated expected profits incurred by the manufacturer based on sales volume and price. The authors' estimation of sales volume included incidence, market share and a deflation factor on price. The research performed in this thesis could inform such a study further by providing a better estimate of health technology diffusion that could be used instead of a time-fixed market share estimate. Willan and Eckermann (2010) and Pezeshk and Gittins (2006) estimated the relationship between the strength of evidence and future number of users of the new treatment. The net benefit resulting from the trial then depended on the strength of evidence. The functional relationships used in each of the analyses were dependent on assumptions that I avoid by eliciting the effect of research on diffusion from experts. I therefore see the potential of this analysis in being relevant to different purposes in industrial decision-making, predominantly in the calculation of profits and in the design of valuable research schemes.

9.4 Strengths and limitations

9.4.1 Strengths

The strengths of this thesis relate mainly to the following points:

1. The use of diffusion theory and an established diffusion model

The use of an established diffusion model based on diffusion theory was the only approach to predicting future implementation or forecasting that was grounded in theory and substantiated by a large evidence base. This body of literature stood in comparison with other forecasting methods that were deemed to be less accurate and did not have the same theoretical foundations or evidence to support them (Goodwin et al., 2014).

2. Methods are technology-specific

Diffusion among technologies is heterogeneous and using pre-existing data on alternative technologies may not be appropriate. I addressed this by eliciting technology-specific diffusion data. In doing so, any factors affecting diffusion that were relevant to EIS in PTB screening could be taken into account by the experts. Furthermore, diffusion determinants relevant to EIS were elicited using semi-structured interviews that allowed participants to freely name anything they deemed relevant and to list the most important drivers to EIS diffusion.

3. Applicability to different technologies

While the methods used to obtain diffusion estimates and identify diffusion drivers took account of the individuality of the technology of interest, these methods are generalisable to many types of different technologies. The modelling frameworks presented in Chapter 4 are generic. The interview guide used for identifying drivers to diffusion would require no adaptation to different settings, except in the introductory part in which background of the technology in question is presented. The elicitation protocol does not need altering any more than the interview guide. It is flexible enough to allow elicitation of any different number of diffusion scenarios conditional on different implementation strategies. It is, however, important to consider the choice of experts for each technology. I included clinicians and one business manager but other professional groups could be relevant, too, for example a representative of the manufacturer. The sample of experts should include people who will be involved in the use, purchasing or sales process of the technology in question. Furthermore, the use of tertiles was deemed difficult by experts. Quartiles are an alternative to tertiles that may be more intuitive.

4. Validity of findings on diffusion determinants and their impact on diffusion

The methodology that included both a qualitative and an elicitation study enabled validity checks through triangulation of results. Results from the triangulation process suggested that no new themes were identified through the expert elicitation. They further showed that, while evidence on EIS and the progesterone treatment effect had been deemed most important diffusion factors by a few experts in the interviews, evidence generation schemes proved to have limited effects on improving implementation. This leads to the conclusion that the qualitative study can identify diffusion factors but it cannot tell us anything about the extent to which these factors affect diffusion. The elicitation study

therefore is indispensable in obtaining diffusion estimates. The elicitation of expert beliefs paired with a diffusion model provides a robust framework that stands in stark contrast to previously used static or hypothetical estimates (Willan and Eckermann, 2010) or extrapolation methods that do not make use of established diffusion models (Whyte et al., 2016).

9.4.2 Limitations

1. Assumption of diffusion being independent of study results

The Research EVSIM and subsequently the EVR were calculated using diffusion estimates that did not account for different study results. The result is that this thesis describes the effects of a reduction of uncertainty on diffusion only. I have outlined an approach how this can be incorporated in Chapters 4, 6 and 7. This approach is entails the following steps:

(i) Propose a research study.

(ii) Define cut-off points for plausible research outputs. I suggested using 25% and 75% quartiles and the middle range, but this could be divided up further.

(iii) Report a range of cost-effectiveness outputs for each of these defined ranges.

(iv) For each of these, elicit a diffusion curve using the approach described in Chapter 6, but conditional on the different research and cost-effectiveness outputs.

(v) Based on the elicited diffusion curves, define a relationship between research study results and diffusion.

(vi) Simulate the study in question, with a large number of simulations.

(vii) Apply the appropriate diffusion curve in each of these simulations, conditional on simulated study outcomes.

(viii) Calculate the Research EVSIM using Equation (4.11).

The potential impact on Research EVSIM results have been explored in Chapter 7. However, this was done without having elicited the appropriate diffusion curves conditional on study results, and these results are therefore only illustrative in nature.

2. Resource and time requirements associated with proposed methods

Incorporating diffusion estimates in health technology assessments entails additional resource and time requirements. These are associated with identifying implementation measures, quantifying implementation levels and modelling both, experience curves and the value of implementation. There are two modelling tasks that are the most time-consuming: the simulation of research and the two-stage model for the experience curve. Simulation of trial data entails a few days of analyst time for conceptualising the statistical model for the planned evidence generation scheme and simulating the data. Methods such as the Sheffield Accelerated Value of Information (SAVI) tool (Strong et al., 2015) make the calculation of the EVSI much simpler than it used to be. Modelling the experience curve requires an extension to commonly used models, which entails modelling future periods. Conceptualising this is not difficult, but a PSA could take significantly longer because of it. In the hypothetical example presented in Chapter 8, which was a simple decision tree model with a small number of parameters, the PSA without the experience curve took approximately one hour to run in MS Excel (on a machine with 8GB memory). With the two-stage experience curve model, the PSA ran for more than 12 hours.

Qualitative and elicitation methods are time-consuming, as they require careful planning, ethics approvals (for the qualitative study), scheduling of interviews and data analysis. The elicitation study, now that its design and analysis are in place, can be conducted relatively quicker than the qualitative study. A few days for adapting the design of the protocol and familiarisation with the software and process are sufficient to prepare for this study and the data analysis only takes a couple of days, too. A qualitative study can be a greater undertaking in that ethics approval or R&D governance approvals in the respective trusts are typically required. The interview guide presented here can be used for all health technologies and study preparation, apart from ethics application, is therefore straightforward. However, data analysis is never standardised in qualitative research. Transcribing interviews and going through the various phases of data analysis can take up weeks of a researcher's time.

The additional resource and time requirements can be mitigated by using the proposed analyses only when the conditions described in Section 9.3 hold. When more diffusion data on health technologies becomes available in the future, using diffusion estimates of analogous technologies or meta-analysed data could be a shortcut to doing an elicitation study. Qualitative research to explore factors influencing diffusion may not be needed if the researcher already has a clear idea on what may be causing low uptake in the future. Searching existing catalogues of diffusion factors could help the researcher in assessing what the relevant diffusion drivers are for different technologies. It just has to be borne in mind that taking these shortcuts potentially results in sacrificing accuracy.

3. Deterministic diffusion estimates used in expected value of research calculations

The calculations of the expected value of implementation were performed using the mean diffusion estimates obtained from the elicitation study. With probabilistic diffusion estimates available, further analysis could entail simulations of the EVII measures that take uncertainty surrounding diffusion estimates into account. This was not done due to time constraints in this study, but is an area that could be explored in further research.

4. Only one case study used

The use of one case study only is a limitation as this means that the range of results that could be obtained by performing EVII and DCEA analyses was not fully explored. The reason for which only one case study was chosen was because this enabled the use of qualitative and elicitation research methods to identify and quantify factors influencing the diffusion of EIS. As mentioned above, the qualitative and elicitation studies were time-consuming and could not have been performed multiple times within the scope of this thesis.

5. Reality-check of s-shaped diffusion curves

It was highlighted before that s-shaped diffusion curves are not always reflective of reality. Heterogeneity of technologies and external influencing factors may cause a diffusion curve to exhibit kinks, plateaus or sudden flattening off (Goodwin et al., 2014). Predicting diffusion with an s-shaped curve is therefore to be understood as an approximation to what will happen in the future, a best guess, resulting from the desire to achieve a sensible trade-off of realism and feasibility. Whether the generated diffusion curves are predictive of actual diffusion will have to be established through applying the method and comparing generated curves with actual diffusion data in the future.

6. Utilisation by clinicians ignored

Lastly, I ignored the uptake levels of utilisation by clinicians, which was estimated to be low (Chapter 6). The value of implementation should incorporate this in further research. I did not research learning effects brought about by increased usage because my focus was on the number of purchases as opposed to utilisation and individual usage. An example of users' learning curves in surgery is the reduction of patient's blood loss that occurs with physicians'

experience (Hopper et al., 2007). With EIS, it is not entirely clear yet, whether there will be learning effects in the sense of improving the predictive ability of EIS, and consequently the QALY gain, by increased usage. The pilot study on the predictive ability of EIS had flagged up that impedance results were sensitive to the pressure exerted by the operator of EIS during measurement. But, with the newly developed transducer-based force gauge, this potential problem and source for a learning curve would be eliminated. I therefore do not think that learning curves will play a significant role in EIS. The other reason for ignoring learning curves was that such analyses would require both, estimates of utilisation and purchases.

9.5 Areas for further research

Areas for further research relate to the limitations that are described above. The main limitation was that diffusion was assumed to be independent from study outcomes. I covered ways of addressing this in the previous section and explored potential results in Section 7.4.4. Future research should cover this and explore a range of results.

Furthermore, performing a PSA on the range of possible results of the dynamic EVII analysis in the EIS case study by using probabilistic diffusion estimates could be a worthwhile study. For more accurate value of implementation results, clinicians' utilisation needs to be incorporated in uptake and an elicitation conducted on the effects of research or other implementation strategies on utilisation. This would also enable a study on clinicians' learning curves.

A further effect of diffusion that is worth studying is the potential for generating more evidence as cumulative use increases, which would increase the expected value of implementation further. For instance, if implementation increased and data on predictive ability was collected, then the decision problem could be revisited at a defined point in time and this data taken into account. This effect can be incorporated in value of information analysis by modelling the expected data collection and adjusting the sample size by uptake in each year, for instance within a recommendation with research scheme. A modelling framework for assessing the value of such schemes has recently been developed (Grimm et al., 2016). An increase in implementation would then increase the expected value of information of such a scheme.

Furthermore, applying the proposed qualitative, elicitation and modelling methods in further case studies could have benefits in validating the methods in other settings and gain greater experience with them. The objective of this would be to obtain a better overview of potential outcomes, uses of the framework and processes required to establish this analysis within technology assessments.

The limitation of time and resource requirements for obtaining diffusion estimates could potentially be addressed by the NHS collecting more data on diffusion of health technologies which could then be analysed further to develop a database of diffusion estimates. If and when more diffusion data becomes available, diffusion data for the available periods could be extrapolated over time for a sample of health technologies using the Bass model to start a database on Bass model parameters within health technologies. This could be useful for health technology assessments and value of implementation analyses in which quick estimates of potential health technology diffusion are needed. The database could be used in two ways to populate such analyses: either by choosing an analogous technology, with or without adaptation, or by using meta-analysed diffusion data. Such a database could also inform analyses on the effects of diffusion factors such as the issue of NICE guidance on diffusion data from the UK health care setting. The caveat of this is that diffusion has been shown to be extremely heterogeneous in the review performed in Chapter 3 and huge amounts of data may be required to obtain meaningful results.

9.6 Conclusion

In conclusion, this thesis has contributed to knowledge by developing methods to predict diffusion of health technologies and to obtain more accurate dynamic estimates of the value of implementation, the value of research and cost-effectiveness. These developments are foreseen to be relevant in the context of health technology assessments of medical devices, diagnostics and drugs, particularly in decisions in which there is low implementation and uncertainty; or those technologies with potential for future price changes conditional on uptake. The proposed methods are also relevant to profit and return on investment calculations performed by manufacturers. Further research is needed to obtain diffusion estimates conditional on different study results.

REFERENCES

ABALLEA, S., CHANCELLOR, J., MARTIN, M., WUTZLER, P., CARRAT, F., GASPARINI, R., TONIOLO-NETO, J., DRUMMOND, M. & WEINSTEIN, M. 2007. The cost-effectiveness of influenza vaccination for people aged 50 to 64 years: an international model. *Value in Health,* 10**,** 98-116.

ADES, A. E., LU, G. & CLAXTON, K. 2004. Expected Value of Sample Information Calculations in Medical Decision Modeling. *Medical Decision Making,* 24**,** 207-227.

ANDRONIS, L. & BARTON, P. 2016. Adjusting estimates of the expected value of information for implementation: theoretical framework and practical application. *Med Decis Making,* 36**,** 296-307.

ANUMBA, D. 2013. *RE: Pilot study on EIS for PTB screening.* Type to GRIMM, S.

ATHERLY, D., DREIBELBIS, R., PARASHAR, U. D., LEVIN, C., WECKER, J. & RHEINGANS, R. D. 2009. Rotavirus vaccination: cost-effectiveness and impact on child mortality in developing countries. *Journal of Infectious Diseases,* 200 Suppl 1**,** S28-38.

BANSBACK, N., ARA, R., WARD, S., ANIS, A. & CHOI, H. 2009. Statin Therapy in Rheumatoid Arthritis: A Cost-Effectiveness and Value-of-Information Analysis. *Pharmacoeconomics,* 27**,** 25-37.

BARNETT, J., VASILEIOU, K., DJEMIL, F., BROOKS, L. & YOUNG, T. 2011. Understanding innovators' experiences of barriers and facilitators in implementation and diffusion of healthcare service innovations: a qualitative study. *BMC Health Services Research,* 11**,** 342.

BASS, F. M. 1969. A new product growth model for consumer durables. *Management Science,* 15**,** 215-227.

BASS, F. M. 1980. The relationship between diffusion rates, experience curves and demand elasticities for consumer durable technological innovations. *The Journal of Business,* 53**,** 51-67.

BASS, F. M., GORDON, K., FERGUSON, T. L. & GITHENS, M. L. 2001. DIRECTV: Forecasting diffusion of a new technology prior to product launch. *Interfaces,* 31**,** S82-S93.

BASS, F. M., KRISHNAN, T. V. & JAIN, D. C. 1994. Why the Bass model fits without decision variables. *Marketing Science,* 13**,** 203-223.

BENNIE, M., GODMAN, B., BISHOP, I. & CAMPBELL, S. 2012. Multiple initiatives continue to enhance the prescribing efficiency for the proton pump inhibitors and statins in Scotland. *Expert Review Pharmacoeconomics Outcomes Research,* 12**,** 125-130.

BERGHELLA, V., BAXTER, J. & HENDRIX, N. 2009. Cervical assessment by ultrasound for preventing preterm delivery. *Cochrane Database Syst Rev.,* 3.

BOJKE, L., CLAXTON, K., BRAVO-VERGEL, Y., SCULPHER, M., PALMER, S. & ABRAMS, K. 2010. Eliciting Distributions to Populate Decision Analytic Models. *Value in Health,* 13**,** 557-564.

BOJKE, L., CLAXTON, K., SCULPHER, M. & PALMER, S. 2009. Characterizing Structural Uncertainty in Decision Analytic Models: A Review and Application of Methods. . *Value in Health,* 12.

BOOTH-CLIBBORN, N., PACKER, C. & STEVENS, A. 2000. Health technology diffusion rates. Statins, coronary stents, and MRI in England. *International Journal of Technology Assessment in Health Care,* 16**,** 781-6.

BOTTOMLEY, P. A. & FILDES, R. 1998. The role of prices in models of innovation diffusion. *Journal of Forecasting,* 17**,** 539-555.

BOYLE, E., POULSEN, G., FIELD, D., KURINCZUK, J., WOLKE, D., ALFIREVIC, Z. & QUIGLEY, M. 2012. Effects of gestational age at birth on health outcomes at 3 and 5 years of age: population based cohort study. *BMJ***,** e896.

BREEZE, P. & BRENNAN, A. 2014. Valuing trial designs from a pharmaceutical perspective using value-based pricing. *HEALTH ECONOMICS,* 24**,** 1468-1482.

BRENNAN, A., CHICK, S. E. & DAVIS, R. 2006. A taxonomy of model structures for economic evaluation of health technologies. *HEALTH ECONOMICS,* 15**,** 1295-1310.

BRENNAN, A. & KHARROUBI, S. A. 2007. Efficient computation of partial expected value of sample information using Bayesian approximation. *Journal of Health Economics,* 26**,** 122-48.

BRIGGS, A. H., CLAXTON, K. & SCULPHER, M. 2006. *Decision Modelling for Health Economic Evaluation,* New York, Oxford University Press.

BRITTEN, N. 2006. Qualitative interviews. *In:* POPE, C. & MAYS, N. (eds.) *Qualitative Research in Health Care.* 3rd ed. Malden (MA): Wiley-Blackwell.

BROWN, A., MEENAN, B. J., DIXON, D., YOUNG, T. P. & BRENNAN, M. 2008. Application of the experience curve to price trends in medical devices: Implications for product development and marketing strategies. *Journal of Medical Marketing: Device, Diagnostic and Pharmaceutical Marketing,* 8**,** 241-255.

BROWN, A., MEENAN, B. J. & YOUNG, T. P. 2007. Marketing Innovation: Medical Device Prices Follow the Experience Curve. *Journal of Medical Marketing,* 7**,** 203-212.

BURR, J. M., MOWATT, G., HERNANDEZ, R., SIDDIQUI, M. A. R., COOK, J., LOURENCO, T., RAMSAY, C., VALE, L., FRASER, C., AZUARA-BLANCO, A., DEEKS, J., CAIRNS, J., WORMALD, R., MCPHERSON, S., RABINDRANATH, K. & GRANT, A. 2007. The clinical effectiveness and cost-effectiveness of screening for open angle glaucoma: a systematic review and economic evaluation. *Health Technology Assessment (Winchester, England),* 11**,** iii-iv.

CAMEJO, R., MCGRATH, C. & HERINGS, R. 2011. A dynamic perspective on pharmaceutical competition, drug development and cost effectiveness. *Health Policy,* 100**,** 18-24.

CARLIN, J. B., JACKSON, T., LANE, L., BISHOP, R. F. & BARNES, G. L. 1999. Cost effectiveness of rotavirus vaccination in Australia. *Australian & New Zealand Journal of Public Health,* 23**,** 611-6.

CARO, J., BRIGGS, A. H., SIEBERT, U. & KUNTZ, K. M. 2012. Modeling Good Research Practices—Overview: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force-1. *Value in Health,* 15**,** 796-803.

CASTANEDA-ORJUELA, C., ALVIS-GUZMAN, N., PATERNINA, A. J. & DE LA HOZ-RESTREPO, F. 2011. Cost-effectiveness of the introduction of the pneumococcal polysaccharide vaccine in elderly Colombian population. *Vaccine,* 29**,** 7644-50.

CHAPMAN, A.-M., TAYLOR, C. A. & GIRLING, A. J. 2014. Are the UK Systems of Innovation and Evaluation of Medical Devices Compatible? The Role of NICE's Medical

Technologies Evaluation Programme (MTEP). *Appl Health Econ Health Policy,* 2014**,** 347-357.

CHEW, D. P., CARTER, R., RANKIN, B., BOYDEN, A. & EGAN, H. 2010. Cost effectiveness of a general practice chronic disease management plan for coronary heart disease in Australia. *Australian Health Review,* 34**,** 162-9.

CHILCOTT, J., BRENNAN, A., BOOTH, A., KARNON, J. & TAPPENDEN, P. 2003. The role of modelling in prioritising and planning clinical trials. *Health Technology Assessment (Winchester, England),* 7**,** iii.

CLAES, C., REINERT, R. R. & VON DER SCHULENBURG, J.-M. G. 2009. Cost effectiveness analysis of heptavalent pneumococcal conjugate vaccine in Germany considering herd immunity effects. *European Journal of Health Economics,* 10**,** 25-38.

CLAXTON, K., BRIGGS, A. H., BUXTON, M., CULYER, A. J., MCCABE, C., WALKER, S. & SCULPHER, M. 2008. Value based pricing for NHS drugs: an opportunity not to be missed? *British Medical Journal,* 336**,** 251-4.

CLAXTON, K., FENWICK, E. & SCULPHER, M. J. 2012. Decision-making with uncertainty: the value of information. *Elgar Companion to Health Economics, 2nd Edition***,** 550-562.

CLAXTON, K. & POSNETT, J. 1996. An economic approach to clinical trial design and research priority-setting. *Health Economics,* 5**,** 513-524.

CLEMEN, R. & WINKLER, R. 1999. Combining probability distributions from experts in risk analysis. *Risk analysis,* 19**,** 187-203.

CLEMENTS, K., BARFIELD, W., AYADI, M. & WILBER, N. 2007. Preterm birth-associated cost of early intervention services: an analysis by gestational age. *Pediatrics***,** 866-874.

CLINICAL AUDIT SUPPORT UNIT 2010. National Audit of Angioplasty Procedures. The NHS Information Centre for health and social care.

COFFIN, P. O., SCOTT, J. D., GOLDEN, M. R. & SULLIVAN, S. D. 2012. Cost-effectiveness and population outcomes of general population screening for hepatitis C. *Clinical Infectious Diseases,* 54**,** 1259-71.

COLBOURN, T., ASSEBURG, C., BOJKE, L., PHILIPS, Z., CLAXTON, K., ADES, A. & GILBERT, R. 2007. Prenatal screening and treatment strategies to prevent group B streptococcal and other bacterial infections in early infancy: cost-effectiveness and expected value of information analyses. *Health Technology Assessment,* 11.

COLLINS, R., FENWICK, E., TROWMAN, R., PERARD, R., NORMAN, G., LIGHT, K., BIRTLE, A., PALMER, S. & RIEMSMA, R. 2007. A systematic review and economic model of the clinical effectiveness and cost-effectiveness of docetaxel in combination with prednisone or prednisolone for the treatment of hormone-refractory metastatic prostate cancer. *Health Technology Assessment (Winchester, England),* 11**,** iii-iv.

COOKE, R. M. 1991. *Experts in uncertainty: Opinion and subjective probability in science,* Oxford, Oxford University Press.

COUDEVILLE, L., VAN RIE, A., GETSIOS, D., CARO, J. J., CREPEY, P. & NGUYEN, V. H. 2009. Adult vaccination strategies for the control of pertussis in the United States: an economic evaluation including the dynamic population effects. *PLoS ONE [Electronic Resource],* 4**,** e6284.

COUPE, V. M. H., DE MELKER, H. E., SNIJDERS, P. J. F., MEIJER, C. J. L. M. & BERKHOF, J. 2009. How to screen for cervical cancer after HPV16/18 vaccination in The Netherlands. *Vaccine,* 27**,** 5111-9.

CUCKLE, H. S., RICHARDSON, G. A., SHELDON, T. A. & QUIRKE, P. 1995. Cost effectiveness of antenatal screening for cystic fibrosis.[Erratum appears in BMJ 1995 Dec 16;311(7020):1608]. *BMJ,* 311**,** 1460-3; discussion 1463-4.

CULLUM, N., DAWSON, D., LANKSHEAR, A., LOWSON, K., MAHON, J., RAYNOR, P., SHELDON, T., WATT, I., WEST, P., WRIGHT, D. & WRIGHT, J. 2004. The evaluation of the dissemination, implementation and impact of NICE guidance. National Coordinating Centre for Research Methodology.

CUNNINGHAM, A. D., PLUMMER, C. J., MCCOMB, J. M., LORD, S. W., CUNNINGHAM, M. W., TOUSSAINT, J. M. & RICKARDS, A. F. 2005. The implantable cardioverter-defibrillator: postcode prescribing in the UK 1998–2002. *Heart,* 91**,** 1280-1283.

DALLAT, M., HUNTER, R., TULLY, M., CAIRNS, K. & KEE, F. 2013. A lesson in business: cost-effectiveness analysis of a novel financial incentive intervention for increasing physical activity in the workplace. *BMC Public Health,* 13**,** 953.

DANESHKHAH, A. & OAKLEY, J. 2010. Eliciting Multivariate Probability Distributions. *In:* BOECKER, K. (ed.) *Rethinking risk measurement and reporting.* London.

DE VRIES, R., KRETZSCHMAR, M., SCHELLEKENS, J. F. P., VERSTEEGH, F. G. A., WESTRA, T. A., ROORD, J. J. & POSTMA, M. J. 2010. Cost-effectiveness of adolescent pertussis vaccination for the Netherlands: using an individual-based dynamic model. *PLoS ONE [Electronic Resource],* 5**,** e13392.

DE WALS, P., NGUYEN, V. H., ERICKSON, L. J., GUAY, M., DRAPEAU, J. & ST-LAURENT, J. 2004. Cost-effectiveness of immunization strategies for the control of serogroup C meningococcal disease. *Vaccine,* 22**,** 1233-40.

DEE, A. & HOWELL, F. 2010. A cost-utility analysis of adding a bivalent or quadrivalent HPV vaccine to the Irish cervical screening programme. *European Journal of Public Health,* 20**,** 213-9.

DEPARTMENT OF HEALTH 2006. Usage of cancer drugs approved by NICE Report of Review undertaken by the National Cancer Director. Department of Health.

DEPARTMENT OF HEALTH 2011. Innovation Health and Wealth, Accelerating Adoption and Diffusion in the NHS. Department of Health, NHS Improvement and Efficiency Directorate, Innovation and Service Improvement.

DODD, J., JONES, L., FLENADY, V., CINCOTTA, R. & CROWTHER, C. 2013. Prenatal administration of progesterone for preventing preterm birth in women considered to be at risk of preterm birth (Review). *Cochrane Database of Systematic Reviews*.

DRUMMOND, M., SCULPHER, M., TORRANCE, G., O'BRIEN, B. & STODDART, G. 2005. *Methods for the Economic Evaluation of Health Care Programmes,* New York, Oxford University Press.

ECKERMANN, S. & PEKARSKY, B. 2014. Can the Real Opportunity Cost Stand Up: Displaced Services, the Straw Man Outside the Room. *Pharmacoeconomics,* 32**,** 319-325.

ECKERMANN, S. & WILLAN, A. R. 2007. Expected value of information and decision making in HTA. *Health Economics,* 16**,** 195-209.

EDDAMA, O., PETROU, S., REGIER, D., NORRIE, J., MACLENNAN, G., MACKENZIE, F. & NORMAN, J. 2010. Study of progesterone for the prevention of preterm birth in twins (STOPPIT): Findings from a trial-based cost-effectiveness analysis. *International Journal of Technology Assessment in Health Care,* 26**,** 141-148.

ESSAT, M., FARIA, R., GOMERSALL, T., GRIMM, S., KEETHARUTH, A., WALKER, S., DIXON, S., PALMER, S. & SCULPHER, M. 2013. Getting cost-effective technologies into practice: the value of implementation - Report on initial scoping review for Phase 1. Sheffield: EEPRU.

FARIA, R., WALKER, S., WHYTE, S., DIXON, S., PALMER, S. & SCULPHER, M. 2014. Getting cost-effective technologies into practice: the value of implementation. An application to novel anticoagulants in the prevention of stroke and systemic embolism. *In:* EEPRU (ed.). University of York: EEPRU.

FENWICK, E., CLAXTON, K. & SCULPHER, M. 2008. The value of implementation and the value of information: combined and uneven development. *Medical Decision Making,* 28**,** 21-32.

FLEURENCE, R. 2007. Setting priorities for research: a practical application of 'payback' and expected value of information. *HEALTH ECONOMICS,* 16**,** 1345-1357.

FONSECA, E., CELIK, E., PARRA, M., SINGH, M. & NICOLAIDES, K. 2007. Progesterone and the Risk of Preterm Birth among Women with a Short Cervix. *N Engl J Med***,** 462-9.

GALE, N., HEATH, G., CAMERON, E., RASHID, S. & REDWOOD, S. 2013. Using the framework method for the analysis of qualitative data in multi-disciplinary health research. *BMC Medical Research Methodology,* 13**,** 1-8.

GARTHWAITE, P., KADANE, J. B. & O'HAGAN, A. 2005. Statistical Methods for Eliciting Probability Distributions. *Journal of the American Statistical Association,* 100**,** 680-700.

GENDERS, T., MEIJBOOM, W., MEIJS, M., SCHUIJF, J., MOLLET, N., WEUSTINK, A., PUGLIESE, F., BAX, J., CRAMER, M., KRESTIN, G., DE FEYTER, P. & HUNINK, M. 2009. CT Coronary Angiography in Patients Suspected of Having Coronary Artery Disease: Decision Making from Various Perspectives in the Face of Uncertainty. *Radiology,* 253**,** 734-744.

GENEST, C. & ZIDEK, J. 1986. Combining probability distributions: a critique and annotated bibliography. *Statistical Science,* 1**,** 114-148.

GEROSKI, P. 2000. Models of technology diffusion. *Research Policy,* 29**,** 603-625.

GILBERT, W., NESBITT, T. & DANIELSEN, B. 2003. The Cost of Prematurity: Quantification by Gestational Age and Birth Weight. *Obstet Gynecol,* 102**,** 488-492.

GIRLING, A., FREEMAN, G., GORDON, J., POOLE-WILSON, P., SCOTT, D. & LILFORD, R. 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.

GOBOK, R., KOSASIH, M., LIM, M. & MA, H. 2009. Forecasting for Biomedical Device Companies: Application of Techniques for a New Neuromonitoring Device. San Jose State University.

GOODWIN, P., MEERAN, S. & DYUSSEKENEVA, K. 2014. The Challenges of Pre-launch Forecasting of Adoption Time Series for New Durable Products. *International Journal of Forecasting,* 30**,** 1082-97.

GRANT, M. & BOOTH, A. 2009. A typology of reviews: an analysis of 14 review types and associated methodologies. *Health Information and Libraries Journal***,** 91-108.

GRAY, A., CLARKE, P. M., WOLSTENHOLME, J. L. & WORDSWORTH, S. 2011. *Applied Methods of Cost-effectiveness Analysis in Healthcare,* Oxford, Oxford University Press.

GREEN, B. B., WANG, C. Y., HORNER, K., CATZ, S., MEENAN, R. T., VERNON, S. W., CARRELL, D., CHUBAK, J., KO, C., LAING, S. & BOGART, A. 2010. Systems of support to increase colorectal cancer screening and follow-up rates (SOS): design, challenges, and baseline characteristics of trial participants. *Contemporary Clinical Trials,* 31**,** 589-603.

GREENHALGH, T., ROBERT, G., MACFARLANE, F., BATE, P. & KYRIAKIDOU, O. 2004. Diffusion of Innovations in Service Organizations: Systematic Review and Recommendations. *The Milbank Quarterly,* 82**,** 581-629.

GRIEBSCH, I., KNOWLES, R., BROWN, J., BULL, C., WREN, C. & DEZATEUX, C. 2007. Comparing the clinical and economic effects of clinical examination, pulse oximetry, and echocardiography in newborn screening for congenital heart defects: A probabilistic cost-effectiveness model and value of information analysis. *International Journal of Technology Assessment in Health Care,* 23**,** 192-204.

GRIMM, S., STRONG, M., BRENNAN, A. & WAILOO, A. 2016. Framework for analysing risk in health technology assessments and its application to managed entry agreements. *In:* NICE DECISION SUPPORT UNIT (ed.). Sheffield: ScHARR, University of Sheffield.

GUEST, G., BUNCE, A. & JOHNSON, L. 2006. How Many Interviews Are Enough? : An Experiment with Data Saturation and Variability. *Field Methods,* 18**,** 59-82.

GURUSAMY, K., WILSON, E., BURROUGHS, A. & DAVIDSON, B. 2012. Intra-Operative vs Pre-Operative Endoscopic Sphincterotomy in Patients with Gallbladder and Common Bile Duct Stones - Cost-Utility and Value-of-Information Analysis. *Applied Health Economics & Health Policy,* 10**,** 15-29.

HALL, P., HULME, C., MCCABE, C., OLUBOYEDE, Y., ROUND, J. & CAMERON, D. 2011. Updated Cost-Effectiveness Analysis of Trastuzumab for Early Breast Cancer: A UK Perspective Considering Duration of Benefit, Long-Term Toxicity and Pattern of Recurrence. *Pharmacoeconomics,* 29**,** 415-432.

HALL, P., MCCABE, C., STEIN, R. & CAMERON, D. 2012. Economic evaluation of genomic test-directed chemotherapy for early stage lymph-node positive breast cancer. *Journal of the National Cancer Institute,* 104.

HAWTHORME, G. & OSBORNE, R. 2005. Population norms and meaningful differences for the Assessment of Quality of Life (AQoL) measure. *Australian and New Zealand Journal of Public Health,* 29**,** 136-142.

HEITMAN, S. J., HILSDEN, R. J., AU, F., DOWDEN, S. & MANNS, B. J. 2010. Colorectal cancer screening for average-risk North Americans: an economic evaluation. *PLoS Medicine / Public Library of Science,* 7**,** e1000370.

HIROTA, R. 1979. Nonlinear partial difference equations v. nonlinear equations reducible to linear equations. *Journal of the Physical Society of Japan,* 46.

HONEST, H., FORBES, C. A., DUREE, K. H., NORMAN, G., DUFFY, S. B., TSOURAPAS, A., ROBERTS, T. E., BARTON, P. M., JOWETT, S. M., HYDE, C. J. & KHAN, K. S. 2009. Screening to prevent spontaneous preterm birth: systematic reviews of accuracy and effectiveness literature with economic modelling. *Health Technology Assessment (Winchester, England),* 13**,** 1-627.

HOOMANS, T., ABRAMS, K. R., AMENT, A. J. H. A., EVERS, S. M. A. A. & SEVERENS, J. L. 2009a. Modeling the Value for Money of Changing Clinical Practice Change A Stochastic Application in Diabetes Care. *Medical Care,* 47**,** 1053-1061.

HOOMANS, T., AMENT, A. J. H. A., EVERS, S. M. A. A. & SEVERENS, J. L. 2011. Implementing guidelines into clinical practice: what is the value? *Journal of Evaluation in Clinical Practice,* 17**,** 606-614.

HOOMANS, T., FENWICK, E., PALMER, S. & CLAXTON, K. 2009b. Value of Information and Value of Implementation: Application of an Analytic Framework to Inform Resource Allocation Decisions in Metastatic Hormone-Refractory Prostate Cancer. *Value in Health,* 12**,** 315-324.

HOOMANS, T., SEIDENFELD, J., BASU, A. & MELTZER, D. 2012. Systematizing the use of value of information analysis in prioritizing systematic reviews. *Methods Research Report.* Rockville, MD 20850: Agency for Healthcare Research and Quality.

HOOMANS, T. M., FENWICK, E. A. L. P., PALMER, S. M. & CLAXTON, K. P. 2009c. Value of Information and Value of Implementation: Application of an Analytic Framework to Inform Resource Allocation Decisions in Metastatic Hormone-Refractory Prostate Cancer. *Value in Health March/April,* 12**,** 315-324.

HOPPER, A., JAMISON, M. & LEWIS, W. 2007. Learning curves in surgical practice. *Postgrad Med J,* 83**,** 777-779.

HOWARD, S. & HARRISON, L. 2005. NICE guidance implementation tracking: data sources, methodology & results. London: ABACUS International.

HOYLE, M. 2010. Historical lifetimes of drugs in England: Application to Value of Information and Cost-Effectiveness Analyses. *Value in Health,* 13**,** 885-892.

HOYLE, M. 2011. Accounting for the drug life cycle and future drug prices in cost-effectiveness analysis. *PharmacoEconomics,* 29**,** 1-15.

HOYLE, M. & ANDERSON, R. 2010. Whose costs and benefits? Why economic evaluations should simulate both prevalent and all future incident patient cohorts. *Medical Decision Making,* 30**,** 426-437.

JENA, A. B. & PHILIPSON, T. J. 2007. Cost-effectiveness as a price control. *Health Affairs,* 26**,** 696-703.

JIANG, Z., BASS, F. M. & ISAACSON BASS, P. 2006. Virtual Bass model and the left-hand data-truncation bias in diffusion of innovation studies. *International Journal of Research in Marketing,* 23**,** 93-106.

JOHNSON, S. R., TOMLINSON, G. A., HAWKER, G. A., GRANTON, J. T. & FELDMAN, B. M. 2010. Methods to elicit beliefs for Bayesian priors: a systematic review. *Journal of Clinical Epidemiology,* 63**,** 355-69.

KIM, T., HONG, J. & KOO, H. 2013. Forecasting diffusion of innovative technology at pre-launch A survey-based method. *Industrial Management & Data Systems,* 113**,** 800-816.

KOTLER, P. & KELLER, K. 2012. *Marketing Management,* New Jersey, Pearson.

KUZEL, A. 1992. Sampling in qualitative inquiry. . *In:* CRABTREE, B. & MILLER, W. (eds.) *Doing qualitative research.* Newbury Park, CA: Sage Publications.

LAWRENCE, K. D. & LAWTON, W. 1981. Applications of Diffusion Models: Some Empirical Results. *In:* WIND, Y., MAHAJAN, V. & CARDOZO, R. (eds.) *New Product Forecasting.* Lexington, MA: Lexington Books.

LEAL, J., WORDSWORTH, S., LEGOOD, R. & BLAIR, E. 2007. Eliciting Expert Opinion for Economic Models: An Applied Example. *Value in Health,* 10**,** 195-203.

LEE, H., KIM, S. G., PARK, H.-W. & KANG, P. 2014. Pre-launch new product demand forecasting using the Bass model: A statistical and machine learning-based approach. *Technological Forecasting & Social Change,* 86**,** 49-64.

LEGARD, R., KEEGAN, J. & WARD, K. 2003. In-depth interviews. *In:* RITCHIE, J. & LEWIS, J. (eds.) *Qualitative Research Practice: A guide for social science students and researchers.* London: Sage Publications Ltd.

LEWIS, J. & RITCHIE, J. 2003. Generalising from qualitative research. *In:* RITCHIE, J. & LEWIS, J. (eds.) *Qualitative research practice: a guide for social sciences students and researchers.* London: Sage Publications.

LILIEN, G. & RANGASWAMY, A. 2004. *Marketing engineering: computer-assisted marketing analysis and planning,* Victoria, Canada, DecisionPro.

MAHAJAN, V., MULLER, E. & BASS, F. M. 1990. New Product Diffusion Models in Marketing: A Review and Directions for Research. *Journal of Marketing,* 54**,** 1-26.

MAHAJAN, V. & SHARMA, S. 1986. A simple algebraic estimation procedure for innovation diffusion models of new product acceptance. *Technological Forecasting & Social Change,* 30**,** 331-345.

MAHESWARAN, H. & BARTON, P. 2012. Intensive case finding and isoniazid preventative therapy in HIV infected individuals in Africa: economic model and value of information analysis. *PLOS ONE* [Online], 7. Available: http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0030457.

MASON, B. W., THOMAS, D. R. H. & SALMON, R. L. 2002. Enhanced surveillance of primary measles mumps and rubella (MMR) immunisation in Wales. *Vaccine,* 20**,** 3635-7.

MASON, J. (ed.) 2002. *Qualitative Researching,* London: SAGE Publications Inc.

MASON, J., FREEMANTLE, N., NAZARETH, I., ECCLES, M., HAINES, A. & DRUMMOND, M. 2001. When is it cost-effective to change the behavior of health professionals? *JAMA,* 286**,** 2988-92.

MAUSKOPF, J. A., SULLIVAN, S. D., ANNEMANS, L., CARO, J., MULLINS, C. D., NUIJTEN, M., ORLEWSKA, E., WATKINS, J. & TRUEMAN, P. 2007. Principles of good practice for Budget Impact Analysis: Report of the ISPOR task force on good research practices - Budget Impact Analysis. *Value in Health,* 10**,** 336-347.

MAYS, N. & POPE, C. 2006. Quality in qualitative health research. *In:* POPE, C. & MAYS, N. (eds.) *Qualitative Research in Health Care.* 3rd ed. Malden (MA): Wiley-Blackwell.

MCCULLAGH, L., WALSH, C. & BARRY, M. 2012. Value-of-information analysis to reduce decision uncertainty associated with the choice of thromboprophylaxis after total hip replacement in the Irish health care setting. *Pharmacoeconomics,* 30**,** 941-959.

MCKENNA, C., WALKER, S., LORGELLY, P., FENWICK, E., BURCH, J., SUEKARRAN, S., BAKHAI, A., WITTE, K., HARDEN, M., WRIGHT, K., WOOLACOTT, N. & PALMER, S. 2012. Cost-effectiveness of aldosterone antagonists for the treatment of post-myocardial infarction heart failure. *Value in Health,* 15**,** 420-428.

MEADE, D. J. 1984. The use of growth curves in forecasting market development - a review and appraisal. *Journal of Forecasting,* 3**,** 429-451.

MEADE, D. J. & ISLAM, T. 2006. Modelling and forecasting the diffusion of innovation - a 25-year review. *International Journal of Forecasting,* 22**,** 519-545.

MOHSENINEJAD, L., VAN BAAL, P., VAN DEN BERG, M., BUSKENS, E. & FEENSTRA, T. 2013. Value of information analysis from a societal perspective: a case study in prevention of major depression. *Value in Health,* 16**,** 490-497.

MORENO, S., NOVIELLI, N. & COOPER, N. 2012. Cost-effectiveness of the implantable HeartMate II left ventricular assist device for patients awaiting heart transplantation. *The Journal of Heart and Lung Transplantation,* 31**,** 450-458.

MORGAN, D. L. 2007. Paradigms Lost and Pragmatism Regained: Methodological Implications of Combining Qualitative and Quantitative Methods. *Journal of Mixed Methods Research,* 1**,** 48-76.

MORRIS, S., BAIO, G., KENDALL, E., VON WAGNER, C., WARDLE, J., ATKIN, W., HALLORAN, S. P., HANDLEY, G., LOGAN, R. F., OBICHERE, A., RAINBOW, S., SMITH, S., SNOWBALL, J. & RAINE, R. 2012. Socioeconomic variation in uptake of colonoscopy following a positive faecal occult blood test result: a retrospective analysis of the NHS Bowel Cancer Screening Programme. *British Journal of Cancer,* 107**,** 765-771.

MORRIS, S., DEVLIN, N. & PARKIN, D. 2007. *Economic Analysis in Health Care*, John Wiley & Sons.

MOSTER, D., LIE, R. & MARKESTAD, T. 2008. Long-term medical and social consequences of preterm birth. *N Engl J Med***,** 262-273.

MURPHY, A., FENWICK, E., TOFF, W., NEILSON, M., BERRY, C., UREN, N., OLDROYD, K. & BRIGGS, A. 2013. Transcatheter aortic valve implantation for severe aortic stenosis: the cost-effectiveness case for inoperable patients in the United Kingdom. *International Journal of Technology Assessment in Health Care,* 29**,** 12-19.

MURPHY, G., CHARLETT, A., JORDAN, L. F., OSNER, N., GILL, O. N. & PARRY, J. V. 2004. HIV incidence appears constant in men who have sex with men despite widespread use of effective antiretroviral therapy. *AIDS,* 18**,** 265-72.

NATIONAL INSTITUTE FOR CARDIOVASCULAR OUTCOMES 2011. National Audit of Percutaneous Coronary Interventional Procedures. UCL.

NICE 2006. NICE implementation uptake report: cox II selective inhibitors. NICE.

NICE 2008. NICE implementation uptake report: statins for the prevention of cardiovascular events. NICE.

NICE 2009a. NICE implementation uptake report: Adefovir dipivoxil and peginterferon alfa-2a for the treatment of chronic hepatitis B. NICE.

NICE 2009b. NICE implementation uptake report: Attention deficit hyperactivity disorder (ADHD), management of ADHD in children, young people and adults. NICE.

NICE 2009c. NICE implementation uptake report: Bipolar disorder. NICE.

NICE 2009d. NICE implementation uptake report: Capecitabine and oxaliplatin in the adjuvant treatment of stage III colon cancer. NICE.

NICE 2009e. NICE implementation uptake report: Donepezil, galantamine, rivastigmine and memantine for the treatment of Alzheimer's disease. NICE.

NICE 2009f. NICE implementation uptake report: Drotrecogin alfa (activated) for severe sepsis. NICE.

NICE 2009g. NICE implementation uptake report: Drugs used in the management of urinary incontinence. NICE.

NICE 2009h. NICE implementation uptake report: drugs used to treat hepatitis C. NICE.

NICE 2009i. NICE implementation uptake report: Ezetimibe for the treatment of primary (heterozygous-familial and non-familial) hypercholesterolemia. NICE.

NICE 2009j. NICE implementation uptake report: Heavy menstrual bleeding. NICE.

NICE 2009k. NICE implementation uptake report: insulin glargine for diabetes (types 1 and 2). NICE.

NICE 2009l. NICE implementation uptake report: Rheumatoid arthritis: the management of rheumatoid arthritis in adults. 2009.

NICE 2009m. NICE implementation uptake report: tacrolimus and pimecrolimus for atopic eczema. NICE.

NICE 2010a. NICE implementation uptake report: Laparoscopic surgery for colorectal cancer. NICE.

NICE 2010b. NICE implementation uptake report: Laparoscopic surgery for inguinal hernia repair. NICE.

NICE 2010c. NICE implementation uptake report: Long-acting reversible contraception (LARC). NICE.

NICE 2010d. NICE implementation uptake report: Omalizumab for severe persistent allergic asthma. NICE.

NICE 2010e. NICE implementation uptake report: Riluzole for the treatment of motor neurone disease. NICE.

NICE 2010f. NICE implementation uptake report: Smoking cessation drugs. NICE.

NICE 2010g. NICE implementation uptake report: Surgical and pharmacological interventions for obesity. NICE.

NICE 2010h. NICE implementation uptake report: Triage, assessment, investigation and early management of head injury in infants, children and adults. NICE.

NICE. 2011a. *Diagnostics Assessment Programme manual* [Online]. Available: https://www.nice.org.uk/MEDIA/DEFAULT/ABOUT/WHAT-WE-DO/NICE-GUIDANCE/NICE-DIAGNOSTICS-GUIDANCE/DIAGNOSTICS-ASSESSMENT-PROGRAMME-MANUAL.PDF [Accessed April 2016.

NICE 2011b. Medical Technologies Evaluation Programme: Methods Guide. London: NICE.

NICE 2011c. NICE implementation uptake report: The management of type 2 diabetes. NICE.

NICE. 2012. *Medical Technologies Evaluation Programme - Methods Guide* [Online]. Available: https://www.nice.org.uk/Media/Default/About/what-we-do/NICE-guidance/NICE-medical-technologies/Medical-technologies-evaluation-programme-methods-guide.pdf [Accessed April 2016.

NICE. 2013. *Guide to the methods of technology appraisal 2013* [Online]. London: NICE. Available: http://www.nice.org.uk/article/pmg9/resources/non-guidance-guide-to-the-methods-of-technology-appraisal-2013-pdf [Accessed April 2016.

NICE. 2015. *Preterm labour and birth* [Online]. Available: http://www.nice.org.uk/guidance/ng25 [Accessed February 2016.

NIHR. 2012. *NIHR - About the HTA programme* [Online]. Available: http://www.hta.ac.uk/about/index.shtml [Accessed 06 June 2012.

NOSYK, B., SHARIF, B., SUN, H., COOPER, C. & ANIS, A. 2011. The Cost-Effectiveness and Value of Information of Three Influenza Vaccination Dosing Strategies for Individuals with Human Immunodeficiency Virus. *PLoS ONE [Electronic Resource],* 6.

O'HAGAN, A., BUCK, C., DANESHKHAH, A., EISER, J., GARTHWAITE, P., JENKINSON, D., OAKLEY, J. & RAKOW, T. (eds.) 2006. *Uncertain Judgements: Eliciting Experts' Probabilities,* Chichester: John Wiley & Sons, 2006.

O'HAGAN, A. & OAKLEY, J. 2010. *SHELF: The Sheffield Elicitation Framework* [Online]. Available: http://www.tonyohagan.co.uk/shelf/ [Accessed 27-09-2012.

O'HAGAN, A. & OAKLEY, J. E. 2004. Probability is perfect, but we can't elicit it perfectly. *Reliability Engineering & System Safety,* 85**,** 239-248.

O'HAGAN, A., STEVENS, J. & MONTMARTIN, J. 2001. Bayesian cost-effectiveness analysis from clinical trial data. *Statistics in Medicine***,** 733-753.

OAKLEY, J., DANESHKHAH, A. & O'HAGAN, A. 2010. Nonparametric Prior Elicitation using the Roulette Method. Available: http://www.tonyohagan.co.uk/academic/pub.html [Accessed 06-11-14].

OAKLEY, J. & O'HAGAN, A. 2007. Uncertainty in Prior Elicitations: a Nonparametric Approach. *Biometrika,* 94**,** 427-441.

ODIBO, A., STAMILIO, D., MACONES, G. & POLSKY, D. 2006. 17alpha-hydroxy-progesterone caproate for the prevention of preterm delivery: A cost-effectiveness analysis. *Obstet Gynecol***,** 492-499.

OSNABRUGGE, R., HEAD, S., GENDERS, T., VAN MIEGHEM, N., DE JAEGERE, P., VAN DER BOON, R., KERKVLIET, M., KALESAN, B., BOGERS, A., KAPPETEIN, P. & HUNINK, M. G. M. 2012. Costs of Transcatheter Versus Surgical Aortic Valve Replacement in Intermediate-Risk Patients. *Ann Thoracic Surgery,* 94**,** 1954-60.

PACKER, C., SIMPSON, S. & STEVENS, A. 2006. International diffusion of new health technologies: a ten country analysis of six health technologies. *International Journal of Technology Assessment in Health Care,* 22**,** 419-428.

PACKER, C., STEVENS, A., COOK, A. & RAFTERY, J. 2004. Diffusion of thrombolysis for acute myocardial infarction from 1981 to 2000 in England: Trend analysis and comparison with need. *International Journal of Technology Assessment in Health Care,* 20**,** 531-536.

PANDOR, A., EASTHAM, J., BEVERLEY, C., CHILCOTT, J. & PAISLEY, S. 2004. Clinical effectiveness and costeffectiveness of neonatal screening for inborn errors of metabolism using tandem mass spectrometry: a systematic review. *Health Technology Assessment,* 8.

PERSONAL SOCIAL SERVICES RESEARCH UNIT. 2015. *Unit Costs of Health and Social Care 2014* [Online]. University of Kent. Available: http://www.pssru.ac.uk/project-pages/unit-costs/2014/ [Accessed July 2015].

PETROU, S. 2003. Economic consequences of preterm birth and low birthweight. *BJOG,* 110**,** 17-23.

PETROU, S., DAKIN, H., ABANGMA, G., BENGE, S. & WILLIAMSON, I. 2010. Cost–Utility Analysis of Topical Intranasal Steroids for Otitis Media with Effusion Based on Evidence from the GNOMETrial. *Value in Health,* 13**,** 543-551.

PETROU, S. & KHAN, K. 2012. Economic costs associated with moderate and late preterm birth: Primary and secondary evidence. *Seminars in Fetal & Neonatal Medicine***,** 170-178.

PEZESHK, H. & GITTINS, J. 2006. Bayesian approach to determine the number of subsequent users of a new treatment. *Statistical Methods in Medical Research,* 15**,** 585-592.

PHAM, B., TEAGUE, L., MAHONEY, J., GOODMAN, L., PAULDEN, M., POSS, J., LI, J., IERACI, L., CACONE, S. & KRAHN, M. 2011. Early Prevention of Pressure Ulcers Among Elderly Patients Admitted Through Emergency Departments: A Cost-effectiveness Analysis. *Annals of Emergency Medicine,* 58**,** 468-478.

PHILIPS, Z., CLAXTON, K. & PALMER, S. 2008. The half-life of truth: What are appropriate time horizons for research decisions? *Medical Decision Making,* 28**,** 287-299.

PHILLIPS, L. D. & BANA E COSTA, C. A. 2007. Transparent prioritisation, budgeting and resource allocation with multi-criteria decision analysis and decision conferencing. *Ann Operations Research,* 154**,** 51-58.

POPE, C., ZIEBLAND, S. & MAYS, N. 2006. Analysing qualitative data. *In:* POPE, C. & MAYS, N. (eds.) *Qualitative Research in Health Care.* 3rd ed. Malden (MA): Wiley-Blackwell.

PURMONEN, T., PANKALAINEN, E., TURUNEN, J., ASSEBURG, C. & MARTIKAINEN, J. 2011. Short-course adjuvant trastuzumab therapy in early stage breast cancer in Finland: Cost-effectiveness and value of information analysis based on the 5-year follow-up results of the inHer Trial. *Acta Oncologica,* 50**,** 344-352.

PUTSIS, W. P. 1998. Parameter variation and new product diffusion. *Journal of Forecasting,* 17**,** 231-257.

QUEZADA, A., BARON-PAPILLON, F., COUDEVILLE, L. & MAGGI, L. 2008. Universal vaccination of children against hepatitis A in Chile: a cost-effectiveness study. *Pan American Journal of Public Health,* 23**,** 303-12.

RAMSEY, S., BLOUGH, D. & SULLIVAN, S. 2008. A Forensic Evaluation of the National Emphysema Treatment Trial Using the Expected Value of Information Approach. *Medical Care,* 46**,** 542-548.

REIN, D. B., HICKS, K. A., WIRTH, K. E., BILLAH, K., FINELLI, L., FIORE, A. E., HOERGER, T. J., BELL, B. P. & ARMSTRONG, G. L. 2007. Cost-effectiveness of routine childhood vaccination for hepatitis A in the United States. *Pediatrics,* 119**,** e12-21.

RETEL, V. & GRIMM, S. January 2013 2013. *RE: Forecasting future developments.* Type to RETEL, V.

RETEL, V. P., BUENO-DE-MESQUITA, J. M., HUMMEL, M. J. M., VAN DE VIJVER, M. J., DOUMA, K. F. L., KARSENBERG, K., VAN DAM, F. S. A. M., VAN KRIMPEN, C., BELLOT, F. E., ROUMEN, R. M. H., LINN, S. C. & VAN HARTEN, W. H. 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.

RETEL, V. P., JOORE, M. A., LINN, S. C., RUTGERS, E. J. & VAN HARTEN, W. H. 2012. Scenario drafting to anticipate future developments in technology assessments. *BioMed Central Research Notes,* 5.

REVILL, P. A. A., WALKER, S. A., MABUGU, T. B., NATHOO, K. J. C., MUGYENYI, P. D., KEKITINWA, A. E., MUNDERI, P. F., BWAKURA-DANGAREMBIZI, M. C., MUSIIME, V. D., BAKEERA-KITAKA, S. E., NAHIRYA-NTEGE, P. F., WALKER, A. S. G., SCULPHER, M. J. A. & GIBB, D. M. G. 2015. Opportunities for improving the efficiency of paediatric HIV treatment programmes. *AIDS,* 29**,** 201-210.

RITCHIE, J. & SPENCER, L. 2002. Qualitative data analysis for applied policy research. *In:* BRYMAN, A. & BURGESS, R. G. (eds.) *Analyzing qualitative data.* SAGE Publications Inc.

RITCHIE, J., SPENCER, L. & O'CONNOR, W. 2003. Carrying out qualitative analysis. *In:* RITCHIE, J. & LEWIS, J. (eds.) *Qualitative Research Practice: A guide for social science students and researchers.* London: SAGE Publications Ltd.

ROBERTS, T. E., ROBINSON, S., BARTON, P. M., BRYAN, S., MCCARTHY, A., MACLEOD, J., EGGER, M. & LOW, N. 2007a. Cost effectiveness of home based population screening for Chlamydia trachomatis in the UK: economic evaluation of chlamydia screening studies (ClaSS) project. *BMJ,* 335**,** 291.

ROBERTS, V. I., ESLER, C. N. & HARPER, W. M. 2007b. What impact have NICE guidelines had on the trends of hip arthroplasty since their publication? *The Journal of Bone and Joint Surgery,* 89**,** 864-867.

ROGERS, E. 2003. *Diffusion of Innovations,* New York, Free Press.

SATOH, D. 2001. A discrete Bass model and its parameter estimation. *Journal of the Operations Research,* 44**,** 1-18.

SCULPHER, M. 2000. Evaluating the cost-effectiveness of interventions designed to increase the utilization of evidence-based guidelines. *Family Practice,* 17**,** S26-S31.

SCULPHER, M. 2008. Subgroups and heterogeneity in cost-effectiveness analysis. *Pharmacoeconomics,* 26**,** 799-806.

SILLUP, G. 1992. Forecasting the adoption of new medical technology using the Bass model. *Journal of Health Care Marketing,* 12**,** 42-51.

SINGH, S., NOSYK, B., SUN, H., CHRISTENSON, J. M., INNES, G. & ANIS, A. H. 2008. Value of information of a clinical prediction rule: informing the efficient use of healthcare and health research resources. *International Journal of Technology Assessment in Health Care,* 24**,** 112-9.

SKINNER, J. & STAIGER, D. 2015. Technology Diffusion and Productivity Growth in Health Care. *Review of Economics and Statistics,* 97**,** 951-964.

SOARES, M., BOJKE, L., DUMVILLE, J., IGLESIAS, C., CULLUM, N. & CLAXTON, K. 2011. Methods to elicit experts' beliefs over uncertain quantities: application to a cost effectiveness transition model of negative pressure wound therapy for severe pressure ulceration. *Statistics in Medicine,* 30**,** 2363-2380.

SOARES, M., WELTON, N., HARRISON, D., PEURA, P., HARI, M., HARVEY, S., MADAN, J., ADES, A., PALMER, S. & ROYAN, K. 2012. An evaluation of the feasibility, cost and value of information of a multicentre randomised controlled trial of intravenous immunoglobulin for sepsis (severe sepsis and septic shock): incorporating a

systematic review, meta-analysis and value of information analysis. *In:* HTA, N. (ed.) *Health Technology Assessment.*

SOETEMAN, D. I., BUSSCHBACH, J. J. V., VERHEUL, R., HOOMANS, T. & KIM, J. J. 2011. Cost-Effective Psychotherapy for Personality Disorders in The Netherlands: The Value of Further Research and Active Implementation. *Value in Health,* 14**,** 229-239.

SONG, Y., LEE, S., ZO, H. & LEE, H. 2015. A hybrid Bass-Markov model for the diffusion of a dual-type device-based telecommunication service: The case of WiBro service in Korea. *Computers & Industrial Engineering,* 79**,** 85-94.

SQUIRES, H. 2014. *A methodological framework for developing the structure of Public Health economic models* [Online]. University of Sheffield. Available: http://etheses.whiterose.ac.uk/5316/ [Accessed July 2015.

STEVENS, J. W. & O'HAGAN, A. 2002. Incorporation of genuine prior information in cost-effectiveness analysis of clinical trial data. *International Journal of Technology Assessment in Health Care,* 18**,** 782-790.

STEVENSON, M., SCOPE, A. & SUTCLIFFE, P. 2010. The Cost-Effectiveness of Group Cognitive Behavioral Therapy Compared with Routine Primary Care forWomen with Postnatal Depression in the UK. *Value in Health,* 13**,** 580-584.

STEVENSON, R., MCCABE, C., PHAROAH, P. D. & COOKE, R. 1996. Cost of care for a geographically determined population of low birth weight infants to age 8-9 years. I. Children without disabilities. *Archives of Disease in Childhood* F114-F117.

STRONG, M., OAKLEY, J. & BRENNAN, A. 2014. Estimating Multiparameter Partial Expected Value of Perfect Information from a Probabilistic Sensitivity Analysis Sample: A Nonparametric Regression Approach. *Med Decision Making***,** 311-326.

STRONG, M., OAKLEY, J., BRENNAN, A. & BREEZE, P. 2015. Estimating the Expected Value of Sample Information Using the Probabilistic Sensitivity Analysis Sample: A Fast Nonparametric Regression-Based Method. *Medical Decision Making,* 35**,** 570-583.

STRONG, M., OAKLEY, J. & CHILCOTT, J. 2012. Managing structural uncertainty in health economic decision analysis: a discrepancy approach. *Journal of the Royal Statistical Society: Applied Statistics,* 61**,** 25-45.

SUGARIS, A. & RELJIN, I. 2012. DVB-T2 technology improvements challenge current strategic planning of ubiquitous media networks. *Eurasip Journal on Wireless Communications and Networking*.

SULLIVAN, W. & PAYNE, K. 2011. The Appropriate Elicitation of Expert Opinion in Economic Models: Making Expert Data Fit for Purpose. *Pharmacoeconomics,* 29**,** 455-459.

SULTAN, F., FARLEY, J. U. & LEHMANN, D. R. 1990. A Meta-Analysis of Applications of Diffusion Models. *Journal of Marketing Research,* 27**,** 70-77.

TERRIS-PRESTHOLT, F., QUAIFE, M. & VICKERMAN, P. 2016. Parameterising User Uptake in Economic Evaluations: The role of discrete choice experiments. *Health Economics,* 25**,** 116-123.

THE HEALTH AND SOCIAL CARE INFORMATION CENTRE 2009. Use of NICE appraised medicines in the NHS in England – Experimental Statistics. NHS.

THE HEALTH AND SOCIAL CARE INFORMATION CENTRE 2012a. NHS Maternity Statistics 2011-12 Summary Report. *Hospital Episode Statistics.* NHS.

THE HEALTH AND SOCIAL CARE INFORMATION CENTRE 2012b. Use of NICE appraised medicines in the NHS in England – Experimental Statistics 2012.

TOWNSEND, C. L., CLIFFE, S. & TOOKEY, P. A. 2006. Uptake of antenatal HIV testing in the United Kingdom: 2000-2003. *Journal of Public Health,* 28**,** 248-52.

TUFFAHA, H., RICKARD, C., WEBSTER, J., MARSH, N., GORDON, L., WALLIS, M. & SCUFFHAM, P. A. 2014. Cost-Effectiveness Analysis of Clinically Indicated Versus Routine Replacement of Peripheral Intravenal Catheters. *Applied Health Econ Health Policy***,** 51-58.

TURNER, S., MORRIS, S., SHERINGHAM, J., HUDSON, E. & FULOP, N. J. 2016. Study protocol: DEcisions in health Care to Introduce or Diffuse innovations using Evidence (DECIDE). *Implementation Science,* 11.

UNICEF. 2013. *Baby-friendly statistics* [Online]. Available: http://www.unicef.org.uk/BabyFriendly/About-Baby-Friendly/Awards/Baby-Friendly-statistics-2015/ [Accessed September 2013].

US DEPARTMENT OF HEALTH AND HUMAN SERVICES & CENTERS FOR DISEASE CONTROL AND PREVENTION 2004. Economic costs associated with mental retardation, cerebral palsy, hearing loss, and vision impairment - United States, 2003. *MMWR Morb Mortal Wkly Rep.*

VAN DEN BULTE, C. & STREMERSCH, S. 2004. Social contagion and income heterogeneity in new product diffusion: a meta-analytic test. *Marketing Science,* 23**,** 530-544.

WAILOO, A., SUTTON, A., COOPER, N., TURNER, D., ABRAMS, K., BRENNAN, A. & NICHOLSON, K. 2008. Cost-Effectiveness and Value of Information Analyses of Neuraminidase Inhibitors for the Treatment of Influenza. *Value in Health,* 11**,** 160-171.

WALKER, S., FARIA, R., WHYTE, S., DIXON, S., PALMER, S. & SCULPHER, M. 2014. Getting cost-effective technologies into practice: the value of implementation. Report on framework for valuing implementation initiatives. *In:* EEPRU (ed.). University of York: EEPRU.

WALKER, S., MASON, A. R., CLAXTON, K., COOKSON, R., FENWICK, E., FLEETCROFT, R. & SCULPHER, M. 2010. Value for money and the Quality and Outcomes Framework in primary care in the UK NHS. *British Journal of General Practice,* 60**,** e213-20.

WALKER, S., SCULPHER, M., CLAXTON, K. & PALMER, S. 2012a. Coverage with evidence development, only in research, risk sharing or patient access scheme? A framework for coverage decisions. *CHE Research Paper.* York: Centre for Health Economics, University of York.

WALKER, S., SCULPHER, M., CLAXTON, K. & PALMER, S. 2012b. Coverage with evidence development, only in research, risk sharing, or patient access scheme? A framework for coverage decisions. *Value in Health,* 15**,** 570-579.

WELTON, N., ADES, A. E., CALDWELL, D. & PETERS, T. J. 2008. Research prioritization based on expected value interventions to increase uptake of breast cancer screening. *J R Statistical Society Series A,* 171**,** 1.

WERNER, E., HAMEL, M., ORZECHOWSKI, K., BERGHELLA, V. & THUNG, S. 2015. Cost-effectiveness of Transvaginal Ultrasound Cervical Length Screening in Singletons without a Prior Preterm Birth: An Update. *American Journal of Obstetrics and Gynecology,* 06.

WERNER, E., HAN, C., PETTKER, C., BUHIMSCHI, C., COPEL, J., FUNAI, E. & THUNG, S. 2011. Universal cervical-length screening to prevent preterm birth: a cost-effectiveness analysis. *Ultrasound Obstetrics Gynecology,* 38**,** 32-37.

WHYTE, S., DIXON, S., FARIA, R., WALKER, S., PALMER, S. & SCULPHER, M. 2014. Getting cost-effective technologies into practice: the value of implementation: An application to B-type natriuretic peptide (BNP) testing in diagnosing chronic heart failure. *In:* EEPRU (ed.). University of Sheffield: EEPRU.

WHYTE, S., DIXON, S., FARIA, R., WALKER, S., PALMER, S., SCULPHER, M. & RADFORD, S. 2016. Estimating the Cost-Effectiveness of Implementation: Is Sufficient Evidence Available? *Value in Health,* 19**,** 138-144.

WIGHT, J., CHILCOTT, J., HOLMES, M. & BREWER, N. 2003. The clinical and cost-effectiveness of pulsatile machine perfusion versus cold storage of kidneys for transplantation retrieved from heartbeating and non-heart-beating donors. *Health Technology Assessment,* 7.

WILLAN, A. R. & ECKERMANN, S. B. 2010. Optimal Clinical Trial Design Using Value Of Information Methods With Imperfect Implementation. *Health Economics,* 19**,** 549-561.

WILSON, E., GURUSAMY, K., GLUUD, C. & DAVIDSON, B. 2010. Cost–utility and value-of-information analysis of early versus delayed laparoscopic cholecystectomy for acute cholecystitis. *British Journal of Surgery,* 97**,** 210-219.

WORLD HEALTH ORGANISATION. 2015. *Purchasing Power Parity 2005* [Online]. WHO. Available: http://www.who.int/choice/costs/ppp/en/ [Accessed July 2015].

WRIGHT, T. P. 1935. Factors affecting the cost of airplanes. *Journal of Aeronautic Science,* 3**,** 275-282.

APPENDIX

A. Appendix to Chapter 2

Table A.1 Review of health economic evaluations and use of implementation estimates

Entry	Author, year	Type of economic evaluation	Condition	Comparators	Reason for implementation estimate	Method of deriving implementation estimates	Impact of implementation on result	Mechanism by which ICER was affected by implementation	Areas for future research relating to implementation in economic evaluation
1	(Aballea et al., 2007)	Budget impact analysis, CEA, PSA	Influenza vaccination	Offering vaccination to 50-64 year old population; to only high risk 50-64 year old population	To calculate incidence of influenza-like symptoms in unvaccinated population from average incidence; deaths avoided; gains in productivity	Data on current uptake available, assumptions made about the uptake of low risk vs. high risk groups based on empirical findings	Uptake affected population costs and effects, higher uptake rates produced lower ICER.	Blended ICER. Age cohort consists of low risk and high risk patients: high risk patients are expected to benefit more from vaccination and uptake among high risk patients is higher - and higher with cohort vaccination than with high risk only.	Potential for evaluating health promotion efforts to increase uptake

2	(Atherly et al., 2009)	Budget impact analysis, CEA for many countries	Rotavirus vaccination	Introducing rotavirus vaccine in GAVI (Vaccine alliance) eligible countries, no vaccine	To explore how demand reacts to different price developments, using the Cennium demand forecasting software with other variables including coverage rates	Vaccine coverage rates from demographic and health surveys and assumption, assuming that highest-risk children have less access to vaccine. Demand in different countries resulted from model that depended on coverage, price, etc.	Population costs and effects and ICER: lower coverage rates lead to larger ICER.	Blended ICER. Larger uptake increases uptake in high-risk groups as well which may affect ICER. No further details provided.	-
3	(Burr et al., 2007)	HTA	Screening for open angle glaucoma	Population screening, current practice	Uptake of community eye test would affect cost-effectiveness of OAG screening.	Historical data and assumption	The higher the uptake rate for no screening (eye test with community optometrist) the higher the ICER for OAG screening.	Blended ICER. When community eye test visits have a high uptake, OAG will be discovered without OAG screening.	Future research on whether uptake rates are realistic for relevant patient groups
4	(Carlin et al., 1999)	CEA	Rotavirus vaccination	Rotavirus vaccination, current practice	To estimate population effects	Historical immunisation rates, 2 scenarios	Not clear: ICER changed in sensitivity analysis but other parameters were varied at the same time.	Not available.	-
5	(Castaneda-Orjuela et al., 2011)	CEA	Pneumococcal polysaccharide vaccine in elderly population	Pneumococcal polysaccharide vaccine, current practice	To estimate population effects	Historical rates of vaccination coverage among elderly population	Uptake affected population effects but whether this affected the ICER is not clear, as no sensitivity analysis was conducted	Not available.	-

6	(Chew et al., 2010)	CEA	Coronary heart disease (CHD)	General practice-based CHD initiative, current practice	To estimate population effects and costs	Assumption: Uptake was the proportion of patients with CHD visiting a GP, GPs willing to undertake programme, patients participating for costs, and effects only accrue for those patients who complete the full cycle. Estimates based on similar programmes in e.g. diabetes or assumptions (point estimate and range)	Scenario testing with low and high uptake values was conducted and ICER was larger with lower uptake.	Estimates of uptake are different for costs and effects because health benefits are only accrued for full completion which means that costs but no effects accrue for patients who do not complete the programme (30% in the basecase). No further details provided.	Robust prospective data demonstrating the benefits and uptake of programme would be valuable in refining the CE model.
7	(Claes et al., 2009)	CEA	Heptavalent pneumococcal conjugate vaccine, current practice	Heptavalent pneumococcal conjugate vaccine, current practice	To estimate herd immunity effects	Estimate based on US uptake rates and general immunisation rates in Germany for first and second year of life.	Higher uptake rates were tested in scenario analysis and had an effect on ICER.	Herd immunity	-
8	(Coffin et al., 2012)	CEA	Chronic Hepatitis C	1-time US adult population screening + high-risk screening, screening of only high-risk individuals	To estimate population effects	Estimate based on reported uptake rates in screening recommendations.	ICER changed with uptake rates.	Not available. Authors were contacted but no reply was received.	-

9	(Collins et al., 2007)	HTA	Hormone-refractory metastatic prostate cancer	Docetaxel w/ prednisone, other chemotherapy regimens	To conduct value of implementation analysis	Estimate based on currently observed shares of treatments as all treatments are available at present	The EVPIM was smaller with larger implementation of the cost-effective strategy.	Straightforward in EVPIM.	No implementation estimate used for ICER or EVPI
10	(Coudeville et al., 2009)	CEA	Pertussis vaccination	Childhood, childhood+adolescent, childhood+adolescent+cocoon (parents of newborn), childhood+adolescent+cocoon+1 dose at 40, childhood+adolescent+routine adult	To estimate herd immunity effects	Estimate based on assumption. Different coverage rates were assumed for different strategies.	One strategy was always dominant independently of vaccine coverage rates.	Herd immunity	Limitation that changing coverage rates over time were not evaluated.
11	(Coupe et al., 2009)	CEA	Prevention of cervical cancer in the Netherlands	HPV vaccination for girls aged 12 and screening, just screening, just vaccination, do nothing	To estimate population effects	Vaccination coverage was based on assumption. Not varied or explored in sensitivity analysis. Screening compliance was based on assumption and varied in scenario analysis.	To estimate the number of cervical cancer cases and deaths. Effect of vaccination coverage on ICER not shown. Change of ICER with screening compliance was only small, lower uptake of screening resulted in higher ICER for vaccination and	Assumption that added vaccination reduced screening compliance - lower screening compliance increased ICER of adding vaccination.	Conservative coverage rate assumed compared to data from Dutch national immunization programme.

							screening.		
12	(Cuckle et al., 1995)	CEA	Cystic fibrosis screening	Sequential carrier testing, couple carrier testing	To estimate the cost per affected pregnancy detected	Range of uptake rates obtained from pilot study reports.	Costs per affected pregnancy detected varied with different uptake rates in both screening strategies.	Not available.	-
13	(de Vries et al., 2010)	CEA	Pertussis vaccination	Universal adolescence pertussis booster vaccination, current practice	To estimate cost-effectiveness	Assumed that coverage rate is the same in adolescents as in infants which was available.	Impact of vaccine coverage on result was not discussed	Not available.	-
14	(De Wals et al., 2004)	CEA	Serogroup C meningococcal disease	Mass immunization campaign, routine vaccination 1 or 3 dose, no vaccine	To obtain population effects and costs	Based on point estimates found in literature.	Not available.	Not available.	-
15	(Dee and Howell, 2010)	CEA	HPV infection and cervical cancer in Ireland	Quadrivalent, bivalent or no HPV vaccine additional to cervical cancer screening	To estimate effects of vaccine uptake on use of services provided to those with genital warts, abnormal smear tests and cervical cancer	Vaccination coverage and screening uptake rate based on assumption and vaccination coverage rate was varied in sensitivity analysis.	ICER is influenced by vaccination coverage rates: the higher the coverage rate the lower the ICER.	Not available - authors have been contacted, no reply.	-

16	(Heitman et al., 2010)	CEA	Colorectal cancer screening	Fecal Occult Blood Test (FOBT), Fecal immunochemical test (FIT), fecal DNA every 3 yrs, flexible sigmoidoscopy, no screening	To estimate cost-effectiveness	Based on uptake reported in RCT on FOBT, assumed to be the same for all technologies, and was varied in sensitivity analysis.	Different uptake rates affected the ICER, but not clear in which direction.	Not available - authors have been contacted, no reply.	Further information on long-term adherence rates for annual stool-based tests is needed.
17	(Quezada et al., 2008)	CEA	Hepatitis A vaccine	Vaccinate all children, no vaccination	To estimate population effects and costs and herd immunity	Scenarios based on assumptions	Cost-savings and effectiveness were affected by different scenarios.	Herd immunity	-
18	(Rein et al., 2007)	CEA	Hepatitis A vaccine	Vaccinate all US children, current policy, no vaccination	To estimate population effects	Based on coverage rates of haemophilus influenzae type b vaccine and from a 2003 survey	ICER changes were not reported, but population effects changed, e.g. deaths averted, productivity losses reduced.	Straightforward for population effects.	-
19	(Retel et al., 2009)	CTA	Breast cancer	Gene expression profiling, current practice	CTA framework demands forecasting future developments to enable, where appropriate, influencing these developments	Diffusion scenarios, using timeline of diffusion phases by Rogers (2003)	To be demonstrated in later publication (see Entry 20)	See Entry 20	-

20	(Retel et al., 2012)	CTA	Breast cancer	Gene expression profiling, current practice	To forecast future developments, specifically diffusion	Scenario drafting, feedback by experts, followed by workshops with experts to choose most likely qualitative scenarios, uptake rates agreed to resemble the uptake rates of clinical guidelines in the Netherlands	ICER changed with uptake. The higher uptake, the lower the ICER.	Blended ICER: co-existence of cost-ineffective and cost-effective technology when uptake was low.	VOI to be conducted, more quantitative estimation of uptake is desirable
21	(Roberts et al., 2007a)	CEA	Chlamydia screening	Opportunistic screening, proactive screening	To examine dynamic effects on cost-effectiveness and consider transmission of infectious diseases	Empirical data from Chlamydia screening studies project	ICER highly sensitive to uptake (one-way sensitivity analysis), higher uptake, lower ICER.	Transmission model considers effects beyond the screened individual.	Incorporate differential uptake estimates according to characteristic; future research should focus on cost-effectiveness of measures to increase uptake

Table A.2 Review of PEVI estimates and the role of uptake in population estimates

Entry	Author (Year)	Condition	Comparators	Population estimate based on:	PEVI uptake-adjusted?	Uptake level or diffusion used	Uptake estimate based on:	Issue of uncertainty around population estimate discussed?
1	McCullagh et al. (2012)	Prophylaxis of venous thromboembolysm after hip replacement	Rivaroxaban, dagibatran etexilate, enoxaparin sodium	THR procedures in acute public hospitals in Ireland and uptake	Yes	Out of those that received prevention, 50% received policy of choice	Assumption	It was acknowledged that population estimates for PEVI were themselves subject to uncertainty, however this was not explored.
2	Welton et al. (2008)	Low uptake of breast cancer screening	Do nothing, send letter, flag in patient record, letter and flag	Women eligible for screening	No	100% (implicitly)	-	-
3	Fleurence (2007)	Two case studies: 1. risk of fracture in osteoporosis patients and 2. pressure ulcers	1. Hormone replacement therapy, bisphosphonates, vitamin D with or without calcium, hip protectors, 2. High-spec foam mattress, alternating pressure mattresses and overlays	Number of patients entering the decision in each year	No	100% (implicitly)	-	The implementation of research results will not automatically follow the logical implications of the cost-effectiveness evidence and should therefore be explored in further in EVI analysis.
4	Pandor et al. (2004)	Neonatal screening for inborn errors of metabolism	Neonatal tandem mass spectrometry, no treatment	Number of neonates per annum	No	100% (implicitly)	-	-
5	Soares et al. (2012)	Severe sepsis and septic shock	Adjuvant intravenous immunoglobulin, current care	Incidence	No	100% (implicitly)	-	-

6	Hall et al. (2012)	Early stage lymph node positive breast cancer	Oncotype DX 21-gene assay directed chemotherapy, chemotherapy for all	Incidence	No	100% (implicitly)	-	-
7	Gurusamy et al. (2012)	Gallbladder and common bile duct stones (CBD)	Intra-operative versus pre-operative endoscopic sphincterotomy (ES)	Number of patients with laparoscopic cholecystectomy with CBD stones and the number of ES performed each year	No	100% (implicitly)	-	-
8	Pham et al. (2011)	Pressure ulcers in elderly patients admitted through emergency departments	Pressure re-distributing foam mattresses, standard hospital mattresses	Elderly admitted emergency department patients in Ontario	No	100% (implicitly)	-	-
9	Purmonen et al. (2011)	Human epidermal growth factor Receptor 2 (HER2)-positive early breast cancer	Adjuvant trastuzumab, conventional treatment after chemotherapy	Number of HER2-positive breast cancer patients	No	100% (implicitly)	-	-
10	Nosyk et al. (2011)	Influenza in patients with human immunodeficiency virus	Three influenza vaccine dosing strategies, previous dosing strategy	Prevalence of HIV positive individuals in Canada	No	100% (implicitly)	-	-
11	Hall et al. (2011)	Human epidermal growth factor Receptor 2 (HER2)-positive early breast cancer	Adjuvant trastuzumab, conventional treatment after chemotherapy	Annual incidence of breast cancer and rate of over-expression of HER2 and uptake	Yes	67%	Estimates of use of chemotherapy with adjuvant trastuzumab from a study	-

12	Petrou et al. (2010)	Otitis media with effusion	Topical intranasal steroids, no treatment	Number of children potentially eligible based on a trial	No	100% (implicitly)	-	-
13	Wilson et al. (2010)	Acute cholecystitis	Early versus delayed laparoscopic cholecystectomy	Number of laparoscopic cholecystectomy per annum	No	100% (implicitly)	-	-
14	Eddama et al. (2010)	Pre-term birth in twins	Progesterone gel, no treatment	Estimated number of twin pregnancies per annum	No	100% (implicitly)	-	-
15	Stevenson et al. (2010)	Post-natal depression (PND)	Group Cognitive Behavioural Therapy, routine care	Annual incidence of PND	No	100% (implicitly)	-	-
16	Genders et al. (2009)	Suspected coronary artery disease	Computer-tomographic coronary angiography prior to conventional angiography, conventional coronary angiography	Incidence	No	100% (implicitly)	-	-
17	Bansback et al. (2009)	Rheumatoid arthritis (RA)	Statin therapy in addition to conventional treatment, conventional treatment	Number of RA patients	No	100% (implicitly)	-	-

18	Ramsey et al. (2008)	Emphysema	Lung-volume reduction surgery, medical treatment	Number of procedures per annum	No	100% (implicitly)	-	Acknowledged that there was uncertainty associated with the number of procedures but did not address this in the model.
19	Wailoo et al. (2008)	Influenza	Amantadine, zanamivir, oseltamivir	Influenza attack rate in healthy population and rate of influenza like illness	No	100% (implicitly)	-	Acknowledged uncertainty associated with the population estimate but did not address this in the model.
20	Singh et al. (2008)	Patients with chest discomfort presenting to emergency department	Early Disposition Prediction Rule, standard care	Number of individuals presenting to emergency departments with chest discomfort each year	No	100% (implicitly)	-	EVPI was reported with different levels of incidence; implementation was not discussed.
21	Griebsch et al. (2007)	Newborn screening for congenital heart defects	Clinical examination, pulse oximetry, echocardiography	Number of newborns	No	100% (implicitly)	-	-
22	Girling et al. (2007)	End-stage heart failure	Left-ventricular assist device implantation vs. optimal medical management	Cases of ESHF per annum	No	100% (implicitly)	-	-

23	Colbourn et al. (2007)	Prevention of group B streptococcal and other bacterial infections in early infancy	Prenatal screening and treatment strategies	UK population	No	100% (implicitly)	-	-
24	Wight et al. (2003)	Preserving kidneys prior to transplantation	Pulsatile machine perfusion, cold storage	Transplant population	No	100% (implicitly)	-	-
25	Maheswaran and Barton (2012)	Tuberculosis in HIV infected individuals	9 different screening strategies in combination with Isoniazid Preventative Therapy	Annual HIV incidence	No	100% (implicitly)	-	-
26	McKenna et al. (2012)	Post-myocardial infarction (MI) heart failure	Eplerenone, spironolactone	Prevalence and incidence of post-MI heart failure	No	100% (implicitly)	-	-
27	Mohseninejad et al. (2013)	Prevention of depression	Opportunistic screening and contact psychotherapy, no screening	Prevalence of subthreshold depression	No	100% (implicitly)	-	-
28	Dallat et al. (2013)	Quality of life and absenteeism from work	Monitoring physical activity at work, not monitoring	All current Northern Ireland employees	No	100% (implicitly)	-	-
29	Murphy et al. (2013)	Severe aortic stenosis	Transcatheter aortic valve implantation, medical management	Annual number of patients ineligible for surgery	No	100% (implicitly)	-	-

Table A.3 Results of the expected value of implementation review

Entry	Authors	Title	Aim	Method	Applied EVIM framework by Fenwick et al. (2008) in a case study?	Developed EVIM framework by Fenwick et al. (2008) further?	Effect of research on implementation considered?	Dynamics of implementation considered?	Future research	Comments
1	(Fenwick et al., 2008)	The Value of Implementation and the Value of Information: Combined and Uneven Development	To present a framework that simultaneously addresses the problem of allocating funds between investments in research and implementation.	Simple four state world where both information and implementation can be either at the current level or "perfect".	Assessed EVPIM, EVPI and EVP in case studies on orlistat for obesity treatment, zanamivir for treatment of influenza and prophylactic extraction of wisdom teeth.	Not applicable.	No.	No.	Relax assumption that information has no effect on implementation and consider both, positive and negative effects. Incorporate uncertainty around the current level of implementation to estimate the EVPPI of it. Evaluate specific implementation measures, which requires estimation of the change in implementation that can be achieved. Application of framework in an ongoing technology appraisal.	Additional research could focus on incorporating the dynamics of implementation in the analysis.

2	(Hoomans et al., 2009b)	Value of Information and Value of Implementation: Application of an Analytic Framework to Inform Resource Allocation Decisions in Metastatic Hormone-Refractor y Prostate Cancer	To develop a framework to inform the decisions of reimbursement, research and investing in implementation strategies.	EVPI and EVPIM analysis and application of framework in metastatic hormone-refractory prostate cancer. Sensitivity analysis to test effect of assumptions on results.	Assessed EVPIM, EVPI and EVP in the case study example.	No.	No.	No.	Incorporate the cost of reversal in the framework. Calculate EVSI and EVSIM.	This highlights the relevance of considering research and implementation recommendations at the time that a reimbursement decision is made but does not evaluate specific research or implementation measures.
3	(Willan and Eckermann , 2010)	Optimal clinical trial design using value of information methods with imperfect implementation	Relax the assumption of perfect implementation for EVSI analysis to inform optimal clinical trial design.	Value of Information Analysis	Applied EVSI with imperfect implementation levels in the design of two trials: the early external cephalic version trial and the CADET-Hp trial.	Allow sample size calculations when imperfect implementation levels are considered in the EVSI analysis. Allow implementation strategies to take a few years. Allow implementation levels other than current or perfect.	Yes. The effect of a research study with one study objective and different sample sizes was assessed on reduction of uncertainty and implementation.	No.	Incorporate multi-stage trial designs in the proposed framework.	Assume functional relationship between the strength of research evidence and (static) implementation rather than using elicited diffusion data. Do not compare different studies in terms of their impact on implementation.

4	(Hoomans et al., 2011)	Implementing guidelines into clinical practice: what is the value?	Inform decision making about guideline and specific implementation strategies using the total net benefit approach	Total net benefit approach to calculate the EVPIM and EVSIM, that incorporates the costs and effects of implementation strategies into the strategy's net benefit.	Assessed EVPIM and EVSIM of specific implementation measures.	No.	No.	No.	None identified that relate to the framework.	The estimates used for prior and posterior implementation were static.
5	(Soeteman et al., 2011)	Cost-effective Psychotherapy for Personality Disorders in the Netherlands: The Value of Further Research and Active Implementation	To assess the societal value of conducting research and implementation of cost-effective psychotherapy for clusters B and C personality disorders	EVPI and EVPIM analysis at population level.	Assessed EVPIM, EVPI and EVP in psychotherapy case study.	No.	No.	No.	Calculate the EVSI and EVSIM.	An application of the four-state framework by Fenwick et al. (2008).

6	(Walker et al., 2014)	Getting cost-effective technologies into practice: the value of implementation. Report on framework on valuing implementation initiatives.	Framework for allowing evaluation of different implementation strategies.	EVSIM analysis and assessing the value against cost of implementation measures.	No.	Yes. The framework was extended to allow for different levels of implementation in each time period.	No.	Changes to implementation over time were considered possible but there was no application of the general framework.	Use of appropriate time horizon. Elicitation of diffusion curves or using diffusion data of reference technologies. Incorporating uncertainty around implementation estimates into the value of implementation analysis.	This is a framework that does not provide detailed guidance on data requirements and possible sources of the required data.
7	(Faria et al., 2014)	Getting cost-effective technologies into practice: the value of implementation. An application to novel anticoagulants in the prevention of stroke and systemic embolism.	Quantify the value of increasing uptake of novel anticoagulants for patients with atrial fibrillation by applying a value of implementation framework developed previously (see Entry 7).	Value of Implementation Analysis.	Assessed implementation measures for their value of actual implementation (EVSIM).	See Entry 7.	No.	Extrapolated from available utilisation rate at present using linear and polynomial regression to yield utilisation rates over future time periods.	Effectiveness over time and costs of implementation measures. Eligible population, current and target utilisation rates. How implementation will change without implementation measures	Did not use an established model of diffusion and the two regression methods therefore differ significantly in terms of the implementation that can be achieved and the time it takes - completely different implementation dynamics could also be possible.
8	(Whyte et al., 2014)	Getting cost-effective technologies into practice: the value of implementation. An application to	Quantify the value of increasing uptake BNP testing by applying the value of implementation	Value of Implementation Analysis.	Assessed implementation measures for their value of actual implementation (EVSIM).	See Entry 7.	No.	Assumed an s-shaped curve for cumulative implementation based on diffusion theory which was fitted to two	Exploring the use of diffusion curves for the estimation of implementation in future periods.	There not much evidence on the effectiveness of implementation initiatives. The first periods of implementation of BNP testing

		B-type natriuretic peptide (BNP) testing in diagnosing chronic heart failure	framework developed previously (see Entry 7).					periods of observed data for natural diffusion. An estimated effect size of the implementation initiative was applied to implementation in future periods.		before 2010 were ignored. The way the s-shaped curve was fitted lacked evidence on the maximum number of attainable adoptions, as 100% was implicitly assumed.
9	(Tuffaha et al., 2014)	Cost-Effectiveness Analysis of Clinically Indicated Versus Routine Replacement of Peripheral Intravenous Catheters	Assessing the cost-effectiveness of clinically indicated versus routine replacement of peripheral intravenous catheters	Cost-effectiveness analysis alongside a clinical trial, EVPI and EVPIM analyses.	Assessed EVPI and EVPIM.	No.	No.	No.	None identified that relate to the framework.	Baseline implementation was assumed to be at 0%.
10	(Revill et al., 2015)	Opportunities for improving the efficiency of paediatric HIV treatment programmes	To assess monitoring methods for HIV-infected children on ART therapy and the continuation of cotrimoxazole treatment when children are stabilised on ART.	Incremental Cost-Effectiveness Analysis and Value of Implementation Analysis	Assessed the maximum value of investing in any implementation measure (e.g. strengthening procurement systems, improving drug supply chains and tracking systems	No.	No.	No.	None identified that relate to the framework.	The method of using the difference in NMBs does not consider the value of specific implementation measures in terms of their effect on implementation nor their respective costs and more analysis is required to establish which IM could most

					paying providers based on provision of cotrimoxazole), using the EVPIM. Used the INMB to determine that value.					cost-effectively improve implementation.
11	(Andronis and Barton, 2016)	Adjusting Estimates of the Expected Value of Information for Implementation: Theoretical Framework and Practical Application	To present a method of calculating the expected value of further research that accounts for the reality of implementation	Extended Fenwick et al.'s 4-state world to a 6-state world including sample information and improved implementation and calculated an "implementation-adjusted" EVSI	Applied the framework in a stylised illustrative case study on non-small lung cancer	See methods	Yes.	Yes.	Use better implementation dynamics	The authors did not calculate the true effect of research on decision uncertainty but solely focused on the effect of research on implementation. Their IA-EVSI is therefore in fact the EVSIM of research. The conclusion from this work, that the EVSI always over-estimates the IA-EVSI when implementation is imperfect, is wrong.

B. Appendix to Chapter 3

Table B.1 Results of the predicting diffusion methods review

Entry #	Authors (Year)	Title	Objectives	Method used to predict diffusion with no available data	Bass model of new product growth used?	If expert judgement, is uncertainty reflected?	Limitation of method	Relevant authors' comments
1	(Bass et al., 2001)	DIRECTV: Forecasting diffusion of a new technology prior to product launch	Forecast diffusion prior to product launch	Guessing by analogy and management decision on the most appropriate analogous product	Yes	Not applicable	Choosing the right analogy is important but little is known about the best way of choosing the analogy	None
2	(Jiang et al., 2006)	Virtual Bass Model and the left-hand data-truncation bias in diffusion of innovation studies	Present a method for dealing with bias caused by left-hand truncation of data used for guessing by analogy: "the virtual Bass model"	Guessing by analogy: the p and q parameter estimates for analogous products are used to forecast the diffusion pattern for the new product	Yes	Not applicable	Left and right truncation of data used	Guessing by analogy is an established method
3	(Sugaris and Reljin, 2012)	DVB-T2 technology improvements challenge current strategic planning of ubiquitous media networks	Estimating profitability of digital terrestrial TV	Guessing by analogy	Yes	Not applicable	None	None

4	(Kim et al., 2013)	Forecasting diffusion of innovative technology at pre-launch: a survey-based method	Propose systematic method for the diffusion of forecasting technology at pre-launch stage	Design of expert survey with seven items that are both familiar to respondents and transformable into parameters of logistic diffusion model by means of algebraic transformation and local search methods	No, logistic model used	No	Use of logistic model instead of the Bass model Deterministic estimates	Guessing by analogy is difficult because it is hard to find truly analogous products, especially in the case of radical innovations. The Bass model is superior to the logistic model in most aspects of demand forecasting, but the logistic model has merit in terms of its simplicity.
5	(Lee et al., 2014)	Pre-launch new product demand forecasting using the Bass model: A statistical and machine learning-based approach	Propose novel approach to pre-launch forecasting of new product demand	Guessing by analogy but also establishing relationship between product attributes and diffusion characteristics through statistical and machine learning-based approaches applied to real-world examples that are based on expert judgment of product attribute values	Yes	No	Requires the maintenance of large databases on product attributes and diffusion data of similar technologies. Requires expert judgement to value product attributes	A large number of analogous products is required for this to work. Authors excluded all products with fewer than 12 years worth of data for lack of accuracy of resulting estimates
6	(Song et al., 2015)	A hybrid Bass–Markov model for the diffusion of a dual-type device-based telecommunication service: the case of WiBro service in Korea	Develop a new approach, the hybrid Bass-Markov model, to forecast demand of a dual-type device-based service	Guessing by analogy for Bass model, based on weighted average of existing services	Yes	Not applicable	None	None

Table B.2 Overview of diffusion curves in cancer

Entry	Condition, population	Ref	Technology	Diffusion measures, proxies	Time period (data available)	Year of first guidance	Data sources
1	**Colon cancer**	(NICE, 2009d)	Capecitabine	Prescription costs	2000-2009	2003	IMS Health Hospital Pharmacy Audit
2	**Colon cancer**	(NICE, 2009d)	Oxaliplatin	Prescription costs	2000-2009	2003	IMS Health Hospital Pharmacy Audit
3	**Colorectal cancer**	(NICE, 2010a)	Laparoscopic surgery	% of colorectal resections performed laparoscopically	2001-2009	2004	Hospital Episode Statistics national data warehouse
4	**Colorectal cancer**	(Cullum et al., 2004)	Laparoscopic surgery	% colorectal cancer patients having laparoscopic surgery	1998-2001	2000	HES data
5	**Colorectal cancer**	(Green et al., 2010)	Laparoscopic surgery	% of surgery done laparoscopically	1998-2007	2000	HES data
6	**Advanced colorectal cancer**	(Howard and Harrison, 2005)	Oxaliplatin	Patients	2001-2003	2002	IMS Health
7	**Advanced colorectal cancer**	(Howard and Harrison, 2005)	Irinotecan	Patients	2001-2003	2002	IMS Health
8	**Non-small lung cancer**	(Cullum et al., 2004)	Docetaxel	Milligrammes	1997-2002	2000	Hospital pharmacies
9	**Non-small lung cancer**	(Cullum et al., 2004)	Gemcitabine	Milligrammes	1997-2002	2000	Hospital pharmacies
10	**Non-small lung cancer**	(Cullum et al., 2004)	Paclitaxel	Milligrammes	1997-2002	2000	Hospital pharmacies
11	**Non-small lung cancer**	(Cullum et al., 2004)	Vinorelbine	Milligrammes	1998-2002	2000	Hospital pharmacies
12	**Advanced ovarian cancer**	(Howard and Harrison, 2005)	2nd line: Topotecan	Patients	2001-2003	2001	IMS Health Oncology
13	**Advanced ovarian cancer**	(Howard and Harrison, 2005)	2nd line: Carboplatin	Patients	2001-2003	2001	IMS Health Oncology
14	**Advanced ovarian cancer**	(Howard and Harrison, 2005)	2nd line: Pegylated liposomal doxorubicin	Patients	2001-2003	2001	IMS Health Oncology

Entry	Condition, population	Ref	Technology	Diffusion measures, proxies	Time period (data available)	Year of first guidance	Data sources
			hydrochloride				
15	**Malignant glioma**	(Howard and Harrison, 2005)	Temozolomide	Patients	2001-2003	2001	IMS Health
16	**Glioblastoma multiforme, malignant glioma**	(The Health and Social Care Information Centre, 2012b)	Temozolomide	NIC	2000-2011	2001	IMS HPAI system
17	**Glioblastoma multiforme, malignant glioma**	(The Health and Social Care Information Centre, 2012b)	Carmustine implants	Number of implants	2004-2012	2007	IMS HPAI system
18	**Chronic lymphocytic leukaemia**	(Howard and Harrison, 2005)	Fludarabine	Patients	2001-2003	2001	IMS Health
19	**Chronic myeloid leukemia**	(Howard and Harrison, 2005)	Imatinib	Patients	2001-2003	2002	IMS Health
20	**Pancreatic cancer**	(Howard and Harrison, 2005)	Gemcitabine	Patients	2001-2003	2001	IMS Oncology
21	**Metastatic pancreatic cancer**	(The Health and Social Care Information Centre, 2012b)	Erlotinib	Cost	2005-2011	2007	IMS HPAI system
22	**Cancer**	(Department of Health, 2006)	Trastuzumab	Est number of mg (000)	2000-2005	2002	IMS Health Hospital Pharmacy Audit data
23	**Early breast cancer, metastatic breast cancer**	(The Health and Social Care Information Centre, 2012b)	Trastuzumab	Cost	2000-2011	2002	IMS HPAI system
24	**Breast cancer (early & advanced)**	(The Health and Social Care Information Centre, 2009)	Trastuzumab	Cost	2000-2009	2002	HPAI

Entry	Condition, population	Ref	Technology	Diffusion measures, proxies	Time period (data available)	Year of first guidance	Data sources
25	**Cancer**	(Department of Health, 2006)	Imatinib	Est number of mg (000)	2001-2005	2002	IMS Health Hospital Pharmacy Audit data
26	**Cancer**	(Department of Health, 2006)	Rituximab	Est number of mg (000)	2000-2005	2002	IMS Health Hospital Pharmacy Audit data
27	**Cancer**	(Department of Health, 2006)	Oxaliplatin	Est number of mg (000)	2000-2005	2002	IMS Health Hospital Pharmacy Audit data
28	**Renal cell carcinoma, stromal gastrointestinal tumours**	(The Health and Social Care Information Centre, 2012b)	Sunitinib	Cost	2006-2011	2009	IMS HPAI system

Table B.3 Overview of diffusion curves in acute conditions

Entry	Condition, population	Reference	Technology	Diffusion measures, proxies	Time period (data available)	Year of first guidance	Data sources
29	**Hepatitis B**	(The Health and Social Care Information Centre, 2009)	Entecavir	Cost	2006-2009	2008	ePACT & HPAI databases
30	**Severe sepsis**	(NICE, 2009f)	Drotrecogin	Prescription costs	2002-2009	2004	IMS Health Hospital Pharmacy Audit and Prescription cost analysis (PCA) system
31	**Severe sepsis**	(The Health and Social Care Information Centre, 2009)	Drotrecogin	Cost	2002-2009	2004	HPAI
32	**Head Injury**	(NICE, 2010h)	MRI scanning	No. of episodes from HES with investigations of head	2006-2009	2007	HES national data
33	**Head Injury**	(NICE, 2010h)	CT scanning	No. of episodes from HES with investigations of head	2006-2009	2007	HES national data
34	**Head Injury**	(NICE, 2010h)	X-ray	No. of episodes from HES with investigations of head	2006-2008	2007	HES national data
35	**Urinary incontinence**	(NICE, 2009g)	Oxybutynin	Items prescribed	2005-2009	2006	ePACT
36	**Acute coronary syndromes**	(Howard and Harrison, 2005)	Tirofiban	Units	1999-2003	2000	IMS Health
37	**Acute coronary syndromes**	(Howard and Harrison, 2005)	Eptifibatide	Units	1999-2003	2000	IMS Health
38	**Hip replacement**	(Howard and Harrison, 2005)	Metal hips	Units	2000-2003	2002	IMS Health
39	**Hip replacement**	(Roberts et al., 2007b)	Uncemented prostheses	% of hip replacement that are uncemented or hybrid prostheses	1990-2005	2000	TRENT Regional Arthroplasty study (Wales)

Entry	Condition, population	Reference	Technology	Diffusion measures, proxies	Time period (data available)	Year of first guidance	Data sources
40	**Hip replacement**	(Roberts et al., 2007b)	Hybrid prostheses	% of hip replacement that are uncemented or hybrid prostheses	1990-2005	2000	TRENT Regional Arthroplasty study (Wales)
41	**Wisdom teeth**	(Cullum et al., 2004)	Wisdom teeth removal	No. extractions	1992-2001	1997	HES, Dental practice board, Scottish morbidity records, Scottish Practitioner Services
42	**Angioplasty**	(Clinical Audit Support Unit, 2010)	Stents	% Percutaneous coronary intervention (PCI) with drug eluting stents	2002-2009	-	Survey conducted by Clinical Audit Support Unit, all centre but 3 submitted data
43	**Angioplasty**	(Clinical Audit Support Unit, 2010)	Arterial access route catheters	% cases using radial arteries	2004-2009	-	Survey conducted by Clinical Audit Support Unit, all centre but 3 submitted data
44	**Percutaneous coronary interventional (PCI) procedures**	(National Institute for Cardiovascular Outcomes, 2011)	Stents	% PCI with drug eluting stents	2002-2010		Survey conducted by Clinical Audit Support Unit, all centre but 3 submitted data
45	**PCI procedures**	(National Institute for Cardiovascular Outcomes, 2011)	Arterial access route catheters	% cases using radial arteries	2004-2010		Survey conducted by Clinical Audit Support Unit, all centre but 3 submitted data
46	**PCI procedures**	(Booth-Clibborn et al., 2000)	Coronary stents	Cumulative proportion of hospitals (%)	6 years		Questionnaires to West Midlands hospitals
47	**Diagnosis (MRI)**	(Booth-Clibborn et al., 2000)	MRI	Cumulative proportion of hospitals (%)	17 years		Structured questionnaires to West Midlands hospitals

Table B.4 Overview of diffusion curves in chronic conditions

Entry	Condition, population	Reference	Technology	Diffusion measures, proxies	Time period (data available)	Year of first guidance	Data sources
48	**Alzheimers**	(NICE, 2009e)	Donepezil	Items prescribed, costs	1998-2008	2005	IMS Health Hospital Pharmacy Audit and Prescription cost analysis (PCA) system
49	**Alzheimers**	(NICE, 2009e)	Galantamine	Items prescribed, costs	1998-2008	2005	IMS Health Hospital Pharmacy Audit and Prescription cost analysis (PCA) system
50	**Alzheimers**	(NICE, 2009e)	Rivastigmine	Items prescribed, costs	1998-2008	2005	IMS Health Hospital Pharmacy Audit and Prescription cost analysis (PCA) system
51	**Alzheimers**	(NICE, 2009e)	Memantine	Items prescribed, costs	1998-2008	2005	IMS Health Hospital Pharmacy Audit and Prescription cost analysis (PCA) system
52	**Alzheimer's**	(The Health and Social Care Information Centre, 2009)	Donepezil	Cost	2000-2009	2007	ePACT & HPAI databases
53	**Alzheimer's**	(The Health and Social Care Information Centre, 2009)	Galantamine	Cost	2000-2009	2007	ePACT & HPAI databases
54	**Alzheimer's**	(The Health and Social Care Information Centre, 2009)	Rivastigmine	Cost	2000-2009	2007	ePACT & HPAI databases
55	**Asthma**	(NICE, 2010d)	Omalizumab	Prescription costs	2006-2010	2007	IMS Health Hospital Pharmacy Audit
56	**Asthma**	(The Health and Social Care Information Centre, 2009)	Omalimuzab	Cost	2005-2008	2007	HPAI
57	**Atopic Eczema**	(NICE, 2009m)	Tacrolimus	Prescription costs	2002-2008	2004	IMS Health Hospital Pharmacy Audit
58	**Atopic Eczema**	(NICE, 2009m)	Pimecrolimus	Prescription costs	2003-2008	2004	IMS Health Hospital Pharmacy Audit

Entry	Condition, population	Reference	Technology	Diffusion measures, proxies	Time period (data available)	Year of first guidance	Data sources
59	**Severe chronic hand ecxema**	(The Health and Social Care Information Centre, 2012b)	Alitretinoin	Cost	2008-2011	2009	PCA & HPAI databases
60	**ADHD**	(NICE, 2009b)	Methylphenidate	Prescription costs (NIC)	1991-2009	2006	Prescription cost analysis (PCA) system
61	**ADHD**	(NICE, 2009b)	Atomoxetine	Prescription costs (NIC)	1991-2009	2006	Prescription cost analysis (PCA) system
62	**ADHD**	(NICE, 2009b)	Dexamfetamine	Prescription costs (NIC)	2003-2009	2006	Prescription cost analysis (PCA) system
63	**Chronic hepatitis B**	(NICE, 2009a)	Adefovir dipivoxil	Prescription costs	2003-2008	2003	IMS Health Hospital Pharmacy Audit
64	**Chronic hepatitis B**	(NICE, 2009a)	Peginterferon alpha-2a	Prescription costs	2002-2008	2003	IMS Health Hospital Pharmacy Audit
65	**Chronic hepatitis B**	(NICE, 2009a)	Lamivudine	Prescription costs	2003-2008	2003	IMS Health Hospital Pharmacy Audit
66	**Hepatitis C**	(NICE, 2009h)	Ribavirin	Prescription costs	2001-2008	2004	IMS Health Hospital Pharmacy Audit
67	**Motor neurone disease**	(NICE, 2010e)	Riluzole	Items prescribed, prescription costs	2000-2009	2001 (<1 yr)	IMS Health Hospital Pharmacy Audit and Prescription cost analysis (PCA) system
68	**Motor neurone disease**	(Howard and Harrison, 2005)	Riluzole	Units	1999-2003	2001	IMS Health
69	**Motor neurone disease**	(The Health and Social Care Information Centre, 2009)	Riluzole	Cost	1996-2009	2001	ePACT & HPAI databases
70	**Amyotrophic lateral sclerosis form of motor neurone disease**	(The Health and Social Care Information Centre, 2012b)	Riluzole	Cost	2000-2011	2001	PCA & HPAI databases
71	**Osteoarthritis, RA**	(NICE, 2006)	Celecoxib	Items prescribed	2000-2006	2001 (<1 yr)	Prescribing Analysis and Cost Tool (PACT)
72	**Osteoarthritis, RA**	(NICE, 2006)	Etodolac	Items prescribed	2000-2006	2001 (<1 yr)	Prescribing Analysis and Cost Tool (PACT)

Entry	Condition, population	Reference	Technology	Diffusion measures, proxies	Time period (data available)	Year of first guidance	Data sources
73	**Osteoarthritis, RA**	(NICE, 2006)	Etoricoxib	Items prescribed	2000-2006	2001 (<1 yr)	Prescribing Analysis and Cost Tool (PACT)
74	**Osteoarthritis, RA**	(NICE, 2006)	Meloxicam	Items prescribed	2000-2006	2001 (<1 yr)	Prescribing Analysis and Cost Tool (PACT)
75	**Osteoarthritis, RA**	(NICE, 2006)	Rofecoxib	Items prescribed	2000-2006	2001 (<1 yr)	Prescribing Analysis and Cost Tool (PACT)
76	**Osteoarthritis, RA**	(NICE, 2006)	Valdecoxib	Items prescribed	2000-2006	2001 (<1 yr)	Prescribing Analysis and Cost Tool (PACT)
77	**Bipolar disorder**	(NICE, 2009c)	Lithium	Items prescribed, prescription costs	2001-2009	2003	PCA data
78	**Rheumatoid arthritis**	(NICE, 2009l)	Sodium-aurothiomalate	Total net ingredient cost (NIC)	2005-2009	2007	ePACT
79	**Rheumatoid arthritis**	(NICE, 2009l)	Hydroxychlorochline sulphate	Total net ingredient cost (NIC)	2005-2009	2007	ePACT
80	**Rheumatoid arthritis**	(NICE, 2009l)	Leflunomide	Total net ingredient cost (NIC)	2005-2009	2007	ePACT
81	**Rheumatoid arthritis**	(NICE, 2009l)	Methotrexate	Total net ingredient cost (NIC)	2005-2009	2007	ePACT
82	**Rheumatoid arthritis**	(NICE, 2009l)	Adalimumab	Total net ingredient cost (NIC)	2005-2009	2007	ePACT
83	**Rheumatoid arthritis**	(NICE, 2009l)	Etanercept	Total net ingredient cost (NIC)	2005-2009	2007	ePACT
84	**Rheumatoid arthritis**	(NICE, 2009l)	Infliximab	Total net ingredient cost (NIC)	2005-2009	2007	ePACT
85	**Rheumatoid arthritis**	(NICE, 2009l)	Anakinra	Total net ingredient cost (NIC)	2002-2009	2007	ePACT
86	**Rheumatoid arthritis**	(NICE, 2009l)	Abatacept	Total net ingredient cost (NIC)	2007-2009	2007	ePACT
87	**Crohn's disease and Rheumatoid arthritis**	(Howard and Harrison, 2005)	Etanercept	Units	1999-2003	2002	IMS Health
88	**Crohn's disease and Rheumatoid arthritis**	(Howard and Harrison, 2005)	Infliximab	Units	1999-2003	2002	IMS Health
89	**Diabetes**	(NICE, 2009k)	Insulin glargine	Items prescribed	2002-2008	2002	ePACT
90	**Diabetes**	(NICE, 2011c)	Metformin hydrochloride	Items prescribed, NIC	2006-2010	2009	ePACT

Entry	Condition, population	Reference	Technology	Diffusion measures, proxies	Time period (data available)	Year of first guidance	Data sources
91	**Diabetes**	(NICE, 2011c)	Chlorpropamide	Items prescribed, NIC	2006-2010	2009	ePACT
92	**Diabetes**	(NICE, 2011c)	Glibenclamide	Items prescribed, NIC	2006-2010	2009	ePACT
93	**Diabetes**	(NICE, 2011c)	Gliclazide	Items prescribed, NIC	2006-2010	2009	ePACT
94	**Diabetes**	(NICE, 2011c)	Glimeprimide	Items prescribed, NIC	2006-2010	2009	ePACT
95	**Diabetes**	(NICE, 2011c)	Glipizide	Items prescribed, NIC	2006-2010	2009	ePACT
96	**Diabetes**	(NICE, 2011c)	Gliquidone	Items prescribed, NIC	2006-2010	2009	ePACT
97	**Diabetes**	(NICE, 2011c)	Tolbutamide	Items prescribed, NIC	2006-2010	2009	ePACT
98	**Diabetes**	(NICE, 2011c)	Pioglitazone	Items prescribed, NIC	2007-2010	2009	ePACT
99	**Diabetes**	(NICE, 2011c)	Rosiglitazone	Items prescribed, NIC	2007-2010	2009	ePACT
100	**Type 2 diabetes**	(The Health and Social Care Information Centre, 2012b)	Pioglitazone	Cost	2000-2011	2001	PCA database
101	**Type 2 diabetes**	(The Health and Social Care Information Centre, 2012b)	Rosiglitazone	Cost	2000-2011	2001	PCA database
102	**Type 2 diabetes**	(Howard and Harrison, 2005)	Sulphonyluria	Patients	2000-2003	2000	IMS Health
103	**Type 2 diabetes**	(Howard and Harrison, 2005)	Metformin	Patients	2000-2003	2000	IMS Health
104	**Type 2 diabetes**	(Howard and Harrison, 2005)	Pioglitazone	Patients	2000-2004	2000	IMS Health
105	**Multiple sclerosis**	(Howard and Harrison, 2005)	Beta-interferon	Units	1999-2003	2001	IMS Health
106	**Multiple sclerosis**	(8)	Natalizumab	Cost	2007-2009	2007	HPAI

Entry	Condition, population	Reference	Technology	Diffusion measures, proxies	Time period (data available)	Year of first guidance	Data sources
107	**Renal failure**	(Howard and Harrison, 2005)	Home dialysis	% of all haemodialysis patients	1997-2002	2002	IMS Health
108	**Renal failure**	(Howard and Harrison, 2005)	Hospital dialysis	% of all haemodialysis patients	1997-2002	2002	IMS Health
109	**Age-related macular degeneration**	(The Health and Social Care Information Centre, 2012b)	Ranibizumab	Cost	2007-2011	2008	IMS HPAI system
110	**AIDS**	(Murphy et al., 2004)	Antiretroviral therapy (ART)	Prevalent diagnosed HIV infections receiving ART (%)	1997-2001		Henry et al (2000) 'National assessment of HIV indections'
111	**HIV**	(Townsend et al., 2006)	Antenatal HIV testing	Routine offer (no. of units), uptake rates in units (%)	2000-2003		Postal survey of unit-based obstetric respondents UK
112	**Several conditions: Proton pumb inhibitors & statins**	(Bennie et al., 2012)	Omeprazole	Defined daily dose per 1000 inhibitants (Scotland), number of tablets	2001-2010	NA	National Health Services Scotland Warehouse
113	**Several conditions: Proton pumb inhibitors & statins**	(Bennie et al., 2012)	Pantoprazole	Defined daily dose per 1000 inhibitants (Scotland), number of tablets	2001-2010	NA	National Health Services Scotland Warehouse
114	**Several conditions: Proton pumb inhibitors & statins**	(Bennie et al., 2012)	Lanzoprazole	Defined daily dose per 1000 inhibitants (Scotland), number of tablets	2001-2010	NA	National Health Services Scotland Warehouse
115	**Several conditions: Proton pumb inhibitors & statins**	(Bennie et al., 2012)	Rabeprazole	Defined daily dose per 1000 inhibitants (Scotland), number of tablets	2001-2010	NA	National Health Services Scotland Warehouse
116	**Several conditions: Proton pumb inhibitors & statins**	(Bennie et al., 2012)	Esomeprazole	Defined daily dose per 1000 inhibitants (Scotland), number of tablets	2001-2010	NA	National Health Services Scotland Warehouse

Entry	Condition, population	Reference	Technology	Diffusion measures, proxies	Time period (data available)	Year of first guidance	Data sources
117	**Several conditions: Proton pumb inhibitors & statins**	(Bennie et al., 2012)	Simvastatin	Defined daily dose per 1000 inhibitants (Scotland), number of tablets	2001-2010	NA	National Health Services Scotland Warehouse

Table B.5. Overview of diffusion curves in conditions that can be acute and chronic

Entry	Condition, population	Reference	Technology	Diffusion measures, proxies	Time period (data available)	Year of first guidance	Data sources
118	**CV events**	(NICE, 2008)	Atorvastatin	Items prescribed	2002-2008	2005	electronic Prescribing Analysis Cost Tool (ePACT)
119	**CV events**	(NICE, 2008)	Fluvastatin	Items prescribed	2002-2008	2005	electronic Prescribing Analysis Cost Tool (ePACT)
120	**CV events**	(NICE, 2008)	Pravastatin	Items prescribed	2002-2008	2005	electronic Prescribing Analysis Cost Tool (ePACT)
121	**CV events**	(NICE, 2008)	Rosuvastatin	Items prescribed	2002-2008	2005	electronic Prescribing Analysis Cost Tool (ePACT)
122	**CV events**	(NICE, 2008)	Simvastatin	Items prescribed	2002-2008	2005	electronic Prescribing Analysis Cost Tool (ePACT)
123	**Inguinal hernia repair**	(NICE, 2010b)	Laparoscopic surgery	% of hernia repairs performed laparoscopically	2001-2009	2004	Hospital Episode Statistics national data warehouse (NHS Information Centre)
124	**Insomnia**	(The Health and Social Care Information Centre, 2009)	Zolpidem	Cost	2000-2009	2004	ePACT & PCA databases
125	**Insomnia**	(The Health and Social Care Information Centre, 2009)	Zoplicone	Cost	2000-2009	2004	ePACT & PCA databases
126	**Insomnia**	(The Health and Social Care Information Centre, 2009)	Zaleplon	Cost	2000-2009	2004	ePACT & PCA databases
127	**Primary hypercholesterolemia**	(NICE, 2009i)	Ezetimibe, combined with simvastatin	Items prescribed	2004-2008	2007	electronic Prescribing Analysis Cost Tool (ePACT)
128	**Hypercholesterolaemia**	(The Health and Social Care Information Centre, 2012b)	Ezetimibe	DDDs	2003-2011	2007	PCA database
129	**Hypercholesterolaemia**	(The Health and Social Care Information	Ezetimibe	Cost	2003-2009	2007	PCA & HPAI databases

Entry	Condition, population	Reference	Technology	Diffusion measures, proxies	Time period (data available)	Year of first guidance	Data sources
		Centre, 2009)					
130	**Patients in need for central venous catheters**	(Howard and Harrison, 2005)	Ultrasound for placing central venous catheters	Units	1999-2004	2002	IMS Health
131	**Depression**	(Howard and Harrison, 2005)	Computerised cognitive behavioural therapy	CCBT placements	1999-2003	2002	IMS Health
132	**Arrhythmias**	(3)	Implantable cardioverter-defibrillator (ICD)	total number of ICDs	1995-2001	2000	British Pacing and Electrophysiology Group (BPEG) Register
133	**Arrhythmias**	(Cunningham et al., 2005)	Implantable cardioverter-defibrillator	Total implants / million population	1998-2002	2000	National Pacemaker Database
134	**Thrombolysis**	(Packer et al., 2004)	Streptokinase	Patient doses purchased	1981-2001		IMS Health England data
135	**Thrombolysis**	(Packer et al., 2004)	Alteplase	Patient doses purchased	1981-2001		IMS Health England data
136	**Thrombolysis**	(Packer et al., 2004)	Reteplase	Patient doses purchased	1981-2001		IMS Health England data
137	**Thrombolysis**	(Packer et al., 2004)	Tenecteplase	Patient doses purchased	1981-2001		IMS Health England data
138	**Thrombolysis**	(Packer et al., 2004)	Anistreplase	Patient doses purchased	1981-2001		IMS Health England data

Table B.6 Overview of diffusion curves in other conditions

Entry	Condition, population	Reference	Technology	Diffusion measures, proxies	Time period (data available)	Year of first guidance	Data sources
139	**Heavy menstrual bleeding**	(NICE, 2009j)	Hysterectomy	No. of patients with at least one prescription	1996-2008	2007	Sample of anonymised GP patient records data using IMS Disease Analyser
140	**Heavy menstrual bleeding**	(NICE, 2009j)	Endometrial ablation	No. of patients with at least one prescription	1996-2008	2007	Sample of anonymised GP patient records data using IMS Disease Analyser
141	**Long-acting reversible contraception**	(NICE, 2010c)	Implants	Items prescribed	2004-2010	2005	Prescription Services Division of NHS Business Service Authority
142	**Long-acting reversible contraception**	(NICE, 2010c)	Intrauterine devices	Items prescribed	2004-2010	2005	Prescription Services Division of NHS Business Service Authority
143	**Long-acting reversible contraception**	(NICE, 2010c)	Intrauterine systems	Items prescribed	2004-2010	2005	Prescription Services Division of NHS Business Service Authority
144	**Obesity**	(NICE, 2010g)	Orlistat	Items prescribed	2000-2010	2006	ePACT
145	**Obesity**	(NICE, 2010g)	Sibutramine	Items prescribed	2001-2010	2006	ePACT
146	**Obesity**	(NICE, 2010g)	Bariatric surgery	Items prescribed	1999-2009	2006	ePACT
147	**Obesity**	(Howard and Harrison, 2005)	Sibutramine	Number patients	2001-2004	2001	IMS Health
148	**Obesity**	(3)	Orlistat	Patient months	1997-2002	2001	PACT and hospital pharmacy data
149	**Obesity**	(The Health and Social Care Information Centre, 2009)	Orlistat	Cost	2000-2009	2001	ePACT & HPAI & PCA databases
150	**Obesity**	(The Health and Social Care Information Centre, 2009)	Sibutramine	Cost	2001-2009	2001	ePACT & HPAI & PCA databases
151	**Obesity**	(The Health and Social Care Information	Rimonabant	Cost	2006-2009	2001	ePACT & HPAI & PCA databases

Entry	Condition, population	Reference	Technology	Diffusion measures, proxies	Time period (data available)	Year of first guidance	Data sources
		Centre, 2009)					
152	**Morbid obesity**	(Howard and Harrison, 2005)	Lap bands surgery	Number	2001-2003	2002	IMS Health and reports from 7 out of 12 trusts
153	**Morbid obesity**	(Howard and Harrison, 2005)	Lap bypass	Number	1997-2003	2002	IMS Health and reports from 7 out of 12 trusts
154	**Smoking cessation**	(NICE, 2010f)	Varenicline	Items prescribed	2006-2009	2007	ePACT
155	**Smoking cessation**	(NICE, 2010f)	Buproprion	Items prescribed	2002-2009	2007	ePACT
156	**Smoking cessation**	(NICE, 2010f)	Nicotine replacement therapy (NRT)	Items prescribed	2002-2009	2007	ePACT
157	**Smoking cessation**	(Howard and Harrison, 2005)	NRT	Patients	2001-2003	2002	IMS Health
158	**Smoking cessation**	(Howard and Harrison, 2005)	Buproprion	Patients	2001-2003	2002	IMS Health
159	**Smoking cessation**	(The Health and Social Care Information Centre, 2012b)	Varenicline	Cost	2006-2011	2007	PCA database
160	**Smoking cessation**	(The Health and Social Care Information Centre, 2009)	Varenicline	Cost	2006-2009	2007	ePACT
161	**Prophylaxis of Rhesus disease**	(Howard and Harrison, 2005)	Anti-D	Units	1999-2003	2002	IMS Health
162	**Growth-related diseases**	(Howard and Harrison, 2005)	Somatropin	Units	1999-2003	2002	IMS Health
163	**Opioid dependence**	(6)	Naltrexone	Cost	2000-2011	2007	PCA & HPAI databases
164	**Measles mumps & rubella**	(Mason et al., 2002)	Immunisation	Uptake % of population	1996-2000		Child Health System by Enhanced Surveillance

Table B.7 Results of the predicting diffusion methods review

Entry #	Authors (Year)	Title	Objectives	Method used to predict diffusion with no available data	Bass model of new product growth used?	If expert judgement, is uncertainty reflected?	Limitation of method	Relevant authors' comments
1	(Bass et al., 2001)	DIRECTV: Forecasting diffusion of a new technology prior to product launch	Forecast diffusion prior to product launch	Guessing by analogy and management decision on the most appropriate analogous product	Yes	Not applicable	Choosing the right analogy is important but little is known about the best way of choosing the analogy	None
2	(Jiang et al., 2006)	Virtual Bass Model and the left-hand data-truncation bias in diffusion of innovation studies	Present a method for dealing with bias caused by left-hand truncation of data used for guessing by analogy: "the virtual Bass model"	Guessing by analogy: the p and q parameter estimates for analogous products are used to forecast the diffusion pattern for the new product	Yes	Not applicable	Left and right truncation of data used	Guessing by analogy is an established method
3	(Sugaris and Reljin, 2012)	DVB-T2 technology improvements challenge current strategic planning of ubiquitous media networks	Estimating profitability of digital terrestrial TV	Guessing by analogy	Yes	Not applicable	None	None

4	(Kim et al., 2013)	Forecasting diffusion of innovative technology at pre-launch: a survey-based method	Propose systematic method for the diffusion of forecasting technology at pre-launch stage	Design of expert survey with seven items that are both familiar to respondents and transformable into parameters of logistic diffusion model by means of algebraic transformation and local search methods	No, logistic model used	No	Use of logistic model instead of the Bass model Deterministic estimates	Guessing by analogy is difficult because it is hard to find truly analogous products, especially in the case of radical innovations. The Bass model is superior to the logistic model in most aspects of demand forecasting, but the logistic model has merit in terms of its simplicity.
5	(Lee et al., 2014)	Pre-launch new product demand forecasting using the Bass model: A statistical and machine learning-based approach	Propose novel approach to pre-launch forecasting of new product demand	Guessing by analogy but also establishing relationship between product attributes and diffusion characteristics through statistical and machine learning-based approaches applied to real-world examples that are based on expert judgment of product attribute values	Yes	No	Requires the maintenance of large databases on product attributes and diffusion data of similar technologies. Requires expert judgement to value product attributes	A large number of analogous products is required for this to work. Authors excluded all products with fewer than 12 years worth of data for lack of accuracy of resulting estimates
6	(Song et al., 2015)	A hybrid Bass–Markov model for the diffusion of a dual-type device-based telecommunication service: the case of WiBro service in Korea	Develop a new approach, the hybrid Bass-Markov model, to forecast demand of a dual-type device-based service	Guessing by analogy for Bass model, based on weighted average of existing services	Yes	Not applicable	None	None

C. Appendix to Chapter 5

C.1. Topic guide

Identifying factors that will influence usage of a new technology for pre-term birth screening in the UK

TOPIC GUIDE

Version 1

1. Aim of interviews

- To understand influencing factors of usage of a new technology for PTB screening in the UK.

2. Introduction

- Introduce yourself
- Explain that this study will inform an economic model for a new technology that may diagnose likelihood of PTB
- Explain that this is part of a PhD study that looks into how usage affects the evaluation of a technology
- Explain that I wish to develop a standard approach by which experts are being asked about the usage of an innovative technology.
- Highlight that you will briefly explain a few important concepts
- Explain the concept of uptake as usage of technology at one point in time
- Explain the concept of diffusion as the usage of a technology over a period of time
    - Show graph on first slide and explain that some technologies may never reach 100% of uptake
    - Explain that there is some natural diffusion and that it may be enhanced by implementation strategies
- Explain that uptake and diffusion may be expressed in % of the total eligible patient population
- Explain that a learning curve refers to learning how to use a technology better over time or with usage
- Explain that the interviews will be conducted with different health care providers, business managers and representatives of the manufacturer in order to understand all perspectives
- Reassure re: confidentiality and anonymity.
    - Confirm that the interview is solely for the use of the researchers
    - The report will pull together findings from all participants in the study and no individual will be identified.
- Remind about length of interview – approximately 45 minutes, maximum of 1 hour

- Introduce tape recorder and explain transcription, data storage and destruction (post publication of findings)
- Check if participant has any questions at all at this stage
- Ask participant to sign consent form to provide written consent for participation
- Thank the person for agreeing to participate

3. General questions on PTB management (different versions for different respondent categories)

- Ask about current way of managing PTB
- Ask about point in time in pregnancy where PTB screening becomes relevant
- Ask about different groups of pregnant women with regards to PTB
  - Ask about identification of higher risk women
- Ask about available treatment of women prone to PTB

4. General questions about EIS for PTB

- Provide information and image on EIS technology – use slides
- Ask about the type of patients EIS could be used for
- Ask about the presence of a learning curve for the health care professional
- Ask about ideas of how EIS could be improved
- Ask about their perception of how pricy this technology may be
- Ask about their thoughts on future price developments

5. Main part (take field notes)

- Ask about any characteristics of EIS that may influence uptake
  - Probe for explanations
  - Probe for other factors

Notes:

- Ask whether they can think of anything that is related to the user that may influence uptake
  - Probe: At health care provider level, e.g. skills, motivation, awareness, knowledge
  - Probe: At purchaser's level
  - Probe: At patient level, e.g. acceptance and beliefs
  - Probe for explanations
  - Probe for other factors

Notes:

- Ask whether they can think of anything related to their organization that may influence uptake
  - Probe for e.g. practicalities
  - Probe for explanations
  - Probe for other factors

Notes:

- Ask whether they can think of any characteristics of the external environment that may influence uptake
  - Probe for e.g. policies
  - Probe for explanations
  - Probe for other factors

Notes:

- Ask whether they can think of anything that could increase uptake?
  - Probe: usability, more evidence, communication between peers
  - Probe for explanations
  - If more evidence, probe what type of evidence may increase uptake
  - Probe for other factors

Notes:

- Ask about anything else at all that may influence uptake
  - Probe for explanations
  - Probe for other factors

Notes:

6. Conclusion

- Repeat all the named factors and ask for those three that would affect uptake most

Notes:

Thank the participant for their time and contribution

C.2. Information sheet

Identifying factors that will influence usage of a new technology for pre-term birth screening in the UK

- Information sheet -

You are being invited to take part in a research project. Before you decide, it is important for you to understand why the research is being done and what it will involve. Please read the following information carefully and discuss it with others if you wish. . If there is anything that is not clear or if you would like more information or share concerns you can do so contacting the researcher or a member of staff from the University of Sheffield **(contact details below)**. Take time to decide whether or not you wish to take part. Thank you for reading this.

What is the project's purpose?

The overall aim of this project is to understand factors that will influence usage of a new technology called Electrical Impedance Spectroscopy (EIS) with the purpose of pre-term birth (PTB) screening in the UK and to quantify future usage with the knowledge of these factors. The purpose of this is to test whether a technology's usage affects evaluation in health economic models.

We would like to talk to professionals who in future may be involved in purchasing, using, managing and/or maintaining EIS for PTB screening.

Why have I been chosen?

You have been asked to take part because you may be involved in either purchasing, using, managing or maintaining EIS for PTB screening.

Do I have to take part?

It is up to you whether or not to take part and there will be no negative consequences for you no matter what your decision is. If you do decide to take part you will be asked to sign a consent form and you can keep this information sheet. You are free to withdraw at any time and without giving a reason.

What will happen to me if I take part?

You will participate in an interview in which you will be asked about:

1. Current diagnosis and management of pre-term birth.
2. Your thoughts about the use of EIS for PTB (description of the technology will be provided at the interview).
3. Potential factors that may influence the use of EIS for PTB.

You may also be asked if you agree to be invited to participate in a study that will be held at a later date to elicit quantitative estimates of the technology's use. Your agreement does not mean

that you have to take part, only that you are interested to hear from us. More information will be provided about this.

How long will the study last?

Your involvement will be for one interview of approximately 45 minutes, maximum 1 hour.

What if I change my mind during the study?

You are free to withdraw from the study at any time. You will not have to give any reasons for your withdrawal.

Are there any risks or disadvantages to taking part in this study?

We do not anticipate that there will be any risks or disadvantages to taking part.

What are the possible benefits of taking part?

The information you give us will help us inform an economic model of EIS for PTB screening that will assess cost-effectiveness and the value of investing in implementation measures, product development or further research. We think you will find your involvement interesting. We hope that in the future this will benefit people who deliver, use or receive EIS, as well as make good use of health care resources.

Will my taking part in the study be kept confidential?

All the information that we collect about you during the course of the research will be kept strictly confidential. You will not be able to be identified in any reports or publications. For data analysis, you will be identified by a code rather than a name and the information will be stored in password-protected computer files at the University of Sheffield which can only be accessed by the researchers. Comments made might be attributed to groups of participants, such as the type of hospital you work at, but your name will not be disclosed.

What will happen to the results of the research project?

The interview findings will inform a subsequent study that will serve to elicit quantitative estimates of the future use of EIS. We aim to publish results of the study in a health care journal and present our findings at professional conferences. If you would like a summary of the results on completion or details of any publications and presentations please let the researcher know through the consent form.

Who has ethically reviewed this study?

This project has obtained University Ethics approval by the School of Health and Related Research's University Ethics committee and has obtained R&D governance approval by the Sheffield Teaching Hospital NHS Foundation Trust and Barnsley Hospital NHS Foundation Trust. The project is currently under review by the Rotherham NHS Foundation Trust review committee.

What if I want to know more?

If you want to know more or share concerns, you can get in touch with the researcher, Sabine Grimm or Dr John Stevens, a member of staff at University of Sheffield.

Contact:

Sabine Grimm
Email: sabine.grimm@sheffield.ac.uk
Phone: 0114 222 6382

ScHARR
Room 1.02, Innovation Centre
217 Portobello Rd
Sheffield S1 4DP

Dr John Stevens
Email : j.w.stevens@sheffield.ac.uk
Phone : 0114 222 6396

HEDS, ScHARR
Regent Court
30 Regent Street
Sheffield S1 4DA

C.3. Consent form

Identifying factors that will influence usage of a new technology for pre-term birth screening in the UK

- Consent form -

Please indicate your agreement with your initials in the right hand-side boxes.

1. I confirm that I have read and understand the information sheet explaining the above research project and I have had the opportunity to ask questions about the project.	
2. I understand that my participation is voluntary and that I am free to withdraw at any time without giving a reason and without there being any negative consequences.	
3. I understand that the information I provide will be securely stored and that access will be restricted to the researchers working on this project. I understand that my name will not be linked to research materials and I will not be identifiable in the report(s) that result from this research.	
4. I understand that, as part of the study, audio recordings of the interview will be made.	
5. I agree to take part in the above research project.	

The following statements are optional. Please indicate your choice by circling it.

6. I agree that I wish to be informed about the study results.	Yes	No
7. I agree that I can be informed about an elicitation exercise related to this study.	Yes	No

Name of participant consent	Name of person taking
Date	Date
Signature	Signature

C.4. Participants' statements

Table C.1 Theme 1: Evidence

Respondents	Predictive ability	Treatment effect	Quality
1	We want a highly predictive tool with good cut-off in impedance readings to distinguish those who will deliver early and those who don't. (p. 6) If it works we could screen all pregnant women at their 20 week scan (p. 3) The ability for it to be used by all pregnant women would be crucial for uptake (p. 4)		The clinical dataset will have to be robust… (p. 5) It'll need validation in multi-centre studies across different settings (p. 6) … how the performance of EIS compares with that of comparators (p. 4)
2	There has to be overwhelming evidence, e.g. likelihood ratios in the 10s and 20s and good reproducibility. (p. 10)	… you could instigate treatment to prevent them from having a PTB. (p. 9) It would have to have very good data to show that the cost of PTB decreased. (p. 11)	If a Cochrane review was done on the number of studies that looked at EIS, the prediction of PT labour and the cost-savings, then it would involve local guideline committees and forums to say we are going to introduce this. (p. 11)

3	If the research behind EIS is powerful enough, it could be made gold standard… (p.16) If the research shows that it could be used in the general population, then this would have implications for when and by whom it is used. (p. 16) ...without trial evidence noone is going to introduce anything. (p. 19)	If you see that it saves lives, and you get results in clinical practice. (p. 19)	
4	...the main thing is whether it works and .... (p. 27) If the evidence came out that you would be screening every pregnant woman, that would have quite big training and logistical issues (p. 30)	Other people are not convinced that the existing treatment works and may need more evidence (p. 24) ...the main thing is … and whether it improves outcomes. (p. 27) Happy healthy babies is a valid endpoint. (p. 28)	The more robust the evidence is, the more likely we are to do it. (p. 31) You don't need this many trials to say it works but you have to have a piece of research that, when you critically appraise it, you can't shoot holes through it. (p.31) So it has to have good external validity (p. 31)
5	Reliability is the most important thing. (p. 35) If you start using it in high risk groups, will it translate to low risk groups, will it get false positives, so I think you need a lot more studies (p. 37)	The big thing is what do you do as a result of a positive test. (p. 37)	Evidence in the shape of randomised-controlled trials, double-blind in fact, is essential. (p. 37)

6	Is there a way to screen everybody so that noone has to have PTB in the future (p. 41) The machine has to be reliable and show specific results... (p. 46)	From a clinical outcomes point of view, is this going to be better? (p. 40) If it gives us better health outcomes, it will be taken up quickly. (p. 42)	I would want this information, that it is much better than what we currently have, to be well scrutinised and publicly used (p. 41)
7	It's got to work. (p. 48) It would have to have the research behind it… (p. 53) You have to have your research first. (p. 54)	The clinical outcomes would have to be very robust. (p. 53)	If the clinical evidence is strong enough, I think it's about engaging the health economy to try and adopt it. (p. 54)
8	I think that this would be the holy grail, that this test could be applied to everybody. (p. 57) Research has to show sensitivity and specificity to make sure that clinicians will use it correctly. (p. 58)	If you can't do anything about the clinical outcome, and save unborn children, it may not be adopted. (p. 60)	You'd have to have very good evidence before it might be widely adopted. (p. 61) It's got to be grounded in good quality research, that was well conducted and peer-reviewed. (p. 63) Actually, there is a morbidity associated with this, putting patients under some degree of stress and telling some of them that they may have a PTB, a psychological morbidity that it could ruin your pregnancy (p. 60)

9	What I would like to see first is that this definitely predicts PTB (p. 66) That's why you've got to do the trials in women who've never had PTB before. (p. 66) It all depends on how good a predictor it was. (p. 68)	...if the treatment stopped PT labour then it would be taken on board. (p. 68) It's a combination of how good the predictive value was and how good the treatment was. (p. 71)	Good research could convince me of whether this could make a difference. It has to be a good-sized study with adequate power. And it needs to be randomised. I'm not bothered about lots of trials as long as it's good quality and it's big enough and not biased in any way by the people who are trying to make it. (p. 72)
10	The biggest thing would be the evidence, that it works, and if there was clear evidence on what the true positives, negatives and false positives, negatives were, that could drive uptake. (pp. 74-75)	If you could prove that there was an intervention to improve the overall outcome, then that would send up the uptake. (pp. 74-75)	

Table C.2 Theme 2: Product and service characteristics

Respondents	Easy to use design	Costs	Marketing
1	What people need is a device that is easily used by operators. (p. 2) It must not look menacing or too big, if it looks sleek and nice it is likely that people would take it up. (p. 2) It should have disposable tips... (p. 2) …should not need charging, whether it's a replaceable or rechargeable battery. (p. 3) It should have wireless technology. (p. 2)	Cost-effectiveness / pricing are vital for uptake as well. (p. 4, 5)	
2	If everyone knew it was there and you could easily use it in clinic setting without other equipment. (p. 11) If that stuff was all disposable, then it would be very simple. (p. 9)	We want the best equipment but might not always get it because the initial outlay is very expensive. (p. 11) If this turned out to be effective, this would be what we are aiming for, as obviously PTB has a huge financial implication and every time you keep a baby out of the special care unit you are saving lots of money. (p. 9) Why should we spend X amount of money, if we are not going to improve our efficiency? (p. 11)	It would have to be minimal outlay or it would be for the company to say we will give you this on trial, completely free, with X many refills for 6 months... (p. 12)

3	It should not need difficult procedures to do it. (p. 17) If the actual design of the device looks slick enough it may make it a winner. (p. 17) It should be low maintenance and something that has to be calibrated every morning, is not going to work. (p. 17)	What they are concerned about is recurrent expenditure (p. 20) If something can save money as well, the business case is a lot easier. (p. 20) If this saves 50 lives this year, they will say buy it. Even if it does not save any money. (p. 20)	
4	The easier it is to do, the more likely we are to do it. (p. 31) I've had issues in the past with devices that were too big and did not fit my hand. (p. 26) It needs to have a disposable cover. (p. 23) It would be great if it had a docking station and you put it in and it downloads data to the computer. (p. 23)	This is probably going to have better uptake, as long as it's not too expensive. (p. 25) The capital costs, the consumable costs… will influence uptake. (p. 26) You would need to budget maintenance into the overall costs. (p. 30) The cost of looking after pre-mature babies in the special care unit is just huge and then they are handicapped and need lifelong input and if we had a simple device that costs just £5,000 that we could use on everybody and predict the ones that deliver before 28 weeks, then we could have some impact. (p. 26)	

5	Ease of use is going to affect uptake of this technology. (p. 34) The design is probably quite important, it should look sleek and easy to use and not big and threatening. (pp. 34-35) If it's easy and quick to use it would influence uptake. (p. 35)	It would be what the initial cost is… It would make a big difference if you had to think very carefully about consumable cost. (p. 33) Servicing costs and replacement are very important for the business case. (p. 36)	
6	Who can do it? Is it something that needs to be done by a consultant or can it be done by a nurse? (p. 41) I'd want to know how robust is the piece of kit. (p. 52) If you've got a probe that will link to a computer by bluetooth or wifi, people think that's wonderful. (p. 43)	When things are expensive they are not taken up… (p. 42) If the kit only lasts a year it can't come in at 100,000 pounds. (p. 42) Affordability comes from two things, how much is it and how long do consumables last? (p. 41) Can this be done by a nurse who is much cheaper than a consultant? (p. 41) Can we prevent that woman from having a baby at 24 weeks and get it for delivery at 34 weeks, that's a hell of a lot of cost-saving. (p. 43)	I think it's how enthusiastic the sales person is, if they believe in what they do and they can put that over to you... (p. 44)

7	Is it easy to use, do they need to refer them to a specialist for it? (p. 51) Portability would influence the decision. (p. 52)	If it's under 5,000 pounds per item it can be a directorate local decision but if it's over it has to go to the capital investment team. (p. 48) ...know the cost outcomes and wider financial impact. (p. 49) … it's about tariff payment catching up with what is happening clinically. If there's a saving in another clinical area it would not account towards our saving. (p. 50) Incentives can be perverse as they encourage you to keep higher cost procedures in order to not lose income because you may not be able to take out actual costs (p. 50)	… negotiate things on maintenance, replacement is key. (p. 51) Some companies will work with you to do that [that the disposable bits are low]. (p. 50)
8		If the disposables were costing 10s of pounds, let's say 30 pounds every time you did it, people would think very carefully. (p. 58) … clinicians are increasingly aware of costs. (p. 59) … there is a cost associated with changing existing practice, it's quite difficult to introduce new technologies in the NHS (p. 62)	

9	As long as the evidence said it was good and it was good in everybody's hands because it was easy to use, then I can see the potential. (p. 66) As long as it was robust and didn't break down at the first attempt of using it, that's what makes things usable. (p. 69) As long as it's quick it should be easily brought into clinical practice. (p. 69)	… would have some probe covers that would need costing in (p. 67) How usable it was would depend on the initial outlay… (p. 67) … you would have to take into account whether you'd have to train someone (p. 67) Cost is the biggest thing at the minute because all trusts are under pressure (p. 68) You'd have to cost it up against the cost of PTB, the cost of steroids, the cost of treatments. (p. 67)	It's how good care the manufacturer gives; that is also things like training packages. (p. 70) It's how good the representative is and how much care they give (p. 70)
10	It should be something readily usable, with not much training required. (p. 75) An ergonomic and easy to use look is the greatest factor, along with the evidence for its use. (p. 75) Ease of use is important, how to store the information. (p. 75)	The biggest thing would be evidence and working out the cost benefit, if you could save 3-5 PTBs per year, the impact that would have on drugs, special care baby units and long-term effects. (p. 75)	I think that the manufacturer would provide training initially because people will ask many questions and not all the information will be available (pp. 75,76)

Table C.3 Theme 3: Organisational set-up

Respondents	Budget	Practicalities	Clinician	Midwives' support
1	Budget is an issue, for example trusts or health funders might decide to buy 4 rather than 8 devices. (p.5)		There is a degree of scepticism in the clinical community. (p. 3)	
2	Everyone is having to make savings of 5% a year. We are having to cut down on this that and the other and then you want to introduce a new technology that's going to cost. (p. 12)	You can't do anything now, without someone being on a training course. (p. 13)	When you have come to a certain level of training or age then it is very difficult to go for something new, it's treated with suspicion. (p. 10-11) If you're ignorant of the data out there you are less likely to use it. (p. 11)	

3		Anything that will reduce the pressure on ultrasound will be welcome. (p. 18)	There are some people that are more enthusiastic than others. (p. 17) There are certain people who are up for innovation and some that aren't. (p. 18)	
4		If you do it instead of scans, then this is much quicker and there's efficiencies to be had. (p. 29)	When new research recommends not to do something anymore, everyone immediately stops doing it, but when you're supposed to do something in addition to what you're already doing, uptake is a bit smaller. (p. 31) A lot of people who do obstetrics in the UK are not obstetricians, they are primarily gynaecologists... And	

			a large chunk of them don't know that evidence exists because their interest is not in obstetrics. (p. 28)	
5		You may need someone to chaperone you and that can be an issue in an antenatal clinic. (p. 36) And who is responsible for maintaining it is quite an issue (p. 36)	There'll be enthusiasts and there'll be people who will say, what am I going to do differently? So I think it takes a while for uptake to go up. (p. 34) We're all a bit resistant to change, you always do the same thing that you've always done and it takes some time to get to it. (p. 35) The fact that junior	You need to buy the midwives' support into it. They are often the ones who see it, they'll suggest why don't you do this test, so I think it's important to engage them. (p. 35)

			staff rotate between units can influence how it's used in practice more than anything else. (p. 36)	
6		If I need to do it to every patient that I see in antenatal clinic, that is going to add 15 minutes and reduce my clinic by a third. (p. 41) If you do it instead of transvaginal scanning, we may not need to buy transvaginal probes anymore. (p. 42)	Where would you be if there is a drug or process that comes along, and I'm probably quite conservative with trying new things. (p. 40)	

7	It is about whether it is capital or revenue budget. We could be totally broke and load the costs at the front end because it's not our directorate's budget, but it's still the trust's budget. (p. 53)	Ultrasound seems to be such a scarce resource that any alternative would be welcome. (p. 50)		
8		It might save you money because you don't have to do other tests, such as ultrasound. (p. 62)	And then you have some trusts, where they are more innovative and keen to innovate and introduce new technologies. (p. 62)	

9	The cost-benefit will influence uptake, and at the minute, this trust is very financially bound. (p. 70)		It's got mostly to do with awareness of how good the technology is. (p. 69)	
10			It depends on the group of doctors and particularly the lead doctor and how they feel about it. (p. 76)	

Table C.4 Theme 4: External factors

Respondents	Guidance	Scientific dissemination	Peer influence	Patient demand	Media
1	Obviously, if it's adopted by NICE that would dramatically improve uptake. (p. 5)		Views of opinion leaders will increase uptake. (p. 5)		

2	It would take local guideline committees to take it up and even better, NICE taking it up. (p. 11)				
3	The essential factor for the diffusion of a device is the reception that the NICE gives to it (p. 16)				

4	If the Royal College or NICE said everyone must have this then it would be difficult for the trust to say No. (p. 30)	Research needs to be publicised and the more it is, the more likely people are to think that this might be helpful. (p. 28)			
5	Local guidelines and national guidelines, in our case, the RCOG Green-top guidelines have a major influence on what is done. (p. 37)		If you have somebody who is well-respected talking at conferences, there is a take-home message. (p. 37)	A lot of bits of equipment are bought because patients in support groups talk about it, ask for it and want to go somewhere where it's available. (p. 37)	

6	If this goes through and it becomes NICE or college accredited that will encourage it to be taken up. (p. 46)		Bad experience elsewhere can bring uptake down (p. 46)	If patients don't like it, it won't get taken up. (p. 41)	User feedback can be a big influence and it can come from women's magazines, soft press stuff and patients. It's getting it out there. (p. 47)
7	When they come with NICE recommendations or are adopted for screening programmes then the funding is held differently. (p. 48)	Clinicians go to conferences, they see all these new toys. (p. 48)			

8	The shape of the uptake curve will alter according to things like evidence by NICE (p. 59)	Publicity is going to be important, that is publicity in the medical fraternity and in the general public (p. 63)		It can become patient-driven, and patients may go to units because there is a degree of choice. (p. 62)	Media is a big influence. (p. 62)
9	If a local network says we're going to use this and you were an outlier then that would influence things. (p. 71)			Women are influenced by other women and they will want to access it… (p. 70)	

10	NICE and the College guidelines would be what sparks a sharp rise in uptake. (p. 77)		The greatest rise in uptake will be from people who just follow the trend (p. 74)		The media can have a strange effect, in front of a paper would do wonders but it also picks up problems (p. 76)

D. Appendix to Chapter 6

D.1. Information sheet for elicitation study

Quantifying factors that will influence usage of a new technology for pre-term birth screening in the UK

- Information sheet -

You are being invited to take part in a research project. Before you decide, it is important for you to understand why the research is being done and what it will involve. Please read the following information carefully and discuss it with others if you wish. If there is anything that is not clear or if you would like more information or share concerns you can do so contacting the researcher or a member of staff from the University of Sheffield **(contact details below)**. Take time to decide whether or not you wish to take part. Thank you for reading this.

What is the project's purpose?

The overall aim of this project is to identify factors that will influence usage of a new technology called Electrical Impedance Spectroscopy (EIS) with the purpose of pre-term birth (PTB) screening in the UK and to quantify future usage with the knowledge of these factors. The purpose of this is to test whether a technology's usage affects evaluation in health economic models.

The relevant factors have already been identified. At this stage, future usage conditional on different levels of these factors will be quantified.

For this exercise, we would like to seek opinions and subjective estimates of professionals who in future may be involved in purchasing, using, managing and/or maintaining EIS for PTB screening.

Why have I been chosen?

You have been asked to take part because you may be involved in either purchasing, using, managing or maintaining EIS for PTB screening.

Do I have to take part?

It is up to you whether or not to take part and there will be no negative consequences for you no matter what your decision is. If you do decide to take part you will be asked to sign a consent form and you can keep this information sheet. You are free to withdraw at any time and without giving a reason.

What will happen to me if I take part?

You will participate in an elicitation exercise that will be conducted at a venue at the University of Sheffield. Elicitation is a widely used method for obtaining experts' judgements. More information about your involvement and the nature of elicitation is provided in the separate 'Elicitation briefing notes' document. It is important that you fill in sections 2 and 4 in the document titled 'Pre-elicitation meeting pro-forma' before the elicitation meeting.

How long will the study last?

Your involvement will be for one elicitation session of approximately three hours.

What if I change my mind during the study?

You are free to withdraw from the study at any time. You will not have to give any reasons for your withdrawal.

Are there any risks or disadvantages to taking part in this study?

We do not anticipate that there will be any risks or disadvantages to taking part.

What are the possible benefits of taking part?

The information you give us will help us inform an economic model of EIS for PTB screening that will assess cost-effectiveness and the value of investing in implementation measures, product development or further research. We think you will find your involvement interesting. We hope that in the future this will benefit people who deliver, use or receive EIS, as well as make good use of health care resources.

Will my taking part in the study be kept confidential?

All the information that we collect about you during the course of the research will be kept strictly confidential. You will not be able to be identified in any reports or publications. For data analysis, you will be identified by a code rather than a name and the information will be stored in password-protected computer files at the University of Sheffield which can only be accessed by the researchers.

To encourage exchange of information in the session it is vital that what is said during the session is treated as confidential and should not be repeated outside the session.

What will happen to the results of the research project?

The quantitative estimates of the future use of EIS will be used in an economic evaluation of EIS. We aim to publish results of the study in a health care journal and present our findings at professional conferences. If you would like a summary of the results on completion or details of any publications and presentations please let the researcher know through the consent form.

Who has ethically reviewed this study?

This project is currently under review by the School of Health and Related Research's University Ethics Review Procedure and under review of the Sheffield Teaching Hospital NHS Foundation Trust R&D governance procedure. [to be changed once approvals are obtained]

What if I want to know more?

If you want to know more or share concerns, you can get in touch with the researcher, Sabine Grimm or Dr John Stevens, a member of staff at University of Sheffield.

Contact:

Sabine Grimm

Email: sabine.grimm@sheffield.ac.uk

Phone: 0114 222 6382

Innovation Centre

217 Portobello Rd

Sheffield S1 4DP

Dr John Stevens

Email : j.w.stevens@sheffield.ac.uk

Phone : 0114 222 6396

HEDS, ScHARR

Regent Court

30 Regent Street

Sheffield S1 4DA

Quantifying uptake determinants of Electrical Impedance Spectroscopy for use in pre-term birth screening in the UK

Elicitation briefing notes

The purpose of the elicitation meeting is to obtain probability distributions to represent your uncertainty about various quantities of interest. These are listed in section three of the attached pro forma.

The elicitation will be conducted following the Sheffield Elicitation Framework (SHELF), based on elicitation practice recommended in O'Hagan et al (2006). You will be given training in the process of elicitation at the start of the meeting, which will include a practice exercise to familiarise you with the procedure.

It is important to note that you will not be asked to provide single estimates of any of these quantities. The elicitation process will instead involve considerations such as what a plausible range of values would be for each unknown quantity, and whether, in your opinion, some values are more likely than others. You may have considerable uncertainty about some of these quantities (though less than that of a lay person). This will not be of concern during the elicitation itself, as the outputs from the elicitation will reflect large uncertainty when it is present.

Due to the subjective nature of elicited probability distributions, it is important to make the elicitation process as transparent as possible. A written record will be kept of the meeting, which will include details of experts present at the meeting, a summary of each expert's relevant expertise, and any declarations of interest. It would be helpful if you could complete sections 2 and 4 in the pro forma. A brief summary will be sufficient for section 4, covering expertise relevant to the parameters listed in section 3.

Please note that declarations of interest are recorded for the purposes of transparency only, and will not be used as grounds for exclusion from the elicitation. It is common for experts to be stakeholders in the wider process.

Suggested relevant evidence is listed in section 5. If you wish to add to this list you may do so. Where appropriate, publications/data listed in this section will be made available at the elicitation meeting.

Reference:

O'Hagan, A., Buck, C. E., Daneshkhah, A., Eiser, J. E., Garthwaite, P. H., Jenkinson, D. J., Oakley, J. E. and Rakow, T. (2006). Uncertain Judgements: Eliciting Expert Probabilities. Chichester: Wiley.

D.3. Pre-elicitation meeting pro forma

Quantifying uptake determinants of Electrical Impedance Spectroscopy for use in pre-term birth screening in the UK

Pre-elicitation meeting pro forma

1) Background	The purpose of this exercise is to obtain estimates of future purchases for Electrical Impedance Spectroscopy (EIS) for use in pre-term birth screening. These estimates will then be used in an economic evaluation of this device that will show future cost-effectiveness and help in setting research priorities.
2) Declarations of interests	[To be completed by the expert. Please identify any personal interest that you might have in the outcome of this elicitation exercise, or in the wider context specified above – whether as an employee, consultant, shareholder or in any other capacity.]
3) Parameter definitions	1. The number of potential purchases of EIS in the NHS England (your expertise of adoption processes in the NHS in your field of work is needed). 2. The number of potential purchases in the first period after introduction of EIS (your expertise of how fast hospitals adopt new technologies in your field of work is needed). 3. The number of years at which the purchase rate peaks (your expertise of how long it takes to reach the majority of adopters in the NHS in your field of work is needed). 4. The same 3 parameters as above assuming that there is more research available (your expertise of impact of research on adoptions in your field of work is needed).
4) Participant's expertise	[To be completed by the expert. Please briefly identify your expertise in relation to the parameters listed above.]

5) Key relevant evidence	[The expert should add details of any key documents and studies that are relevant to the above parameters.]

D.4. Quantifying factors that will influence usage of a new technology for pre-term birth screening in the UK

- Background -

1. Maternity care in the NHS

- Number of hospital births in England: 669,000[1]
- Number of hospital obstetrics units in England: 231[2]
- Out of the above, number of teaching hospitals: ~ 60[3]
- Average number of beds per hospital obstetrics unit: 10[4]
- Average annual number of births per obstetrics unit in England: ~ 2,900 (Range: 10 - 7000)[5]
- Average annual number of births at Jessop wing: ~ 7,000[6]
- Average annual number of births at Rotherham hospital: ~ 2,800[7]

2. Electrical Impedance Spectroscopy for pre-term birth screening

- Electrical Impedance Spectroscopy (EIS) has been used in cervical cancer screening
- Currently in development for use in pre-term birth (PTB) with certain adaptations and improvements
- EIS measures impedance of the cervix
- Use:
  - On pregnant women at high risk based on their history (potentially extendable to women without risk factors but this is NOT in the focus of this exercise)
  - Speculum examination where electrical impedance probe is gently placed on the anterior lip of the cervix
  - Data is captured automatically when button is pressed
- A single-centre clinical experimental study is underway:
  - Recruitment of 250 high risk women (and 250 women without risk factors)
  - Exclusion criteria: history of abnormal cervical smear in previous 3 years, cone biopsy or loop excision of the cervix, women with recent cervical infection or vaginal bleeding, multiple pregnancy or known fetal abnomaly
  - Inclusion criteria: at high risk based on ≥ 1 previous pre-term delivery (< 37 weeks gestation)
  - Cervical Impedance (CI) measurement at 20-22 weeks, repeated at 26-28 weeks

[1] NHS maternity statistics England 2011-2012

[2] www.progress.babyfriendly.org.uk

[3] Wikipedia: Teaching hospitals England http://en.wikipedia.org/wiki/Category:Teaching_hospitals_in_England

[4] NIHR report "Mapping maternity care in England" (2011)

[5] www.birthchoiceuk.com

[6] www.sth.nhs.uk

[7] www.birthchoiceuk.com

- For the purposes of this exercise, we will assume that the predictive values of EIS for pre-term birth measured between 14 and 28 weeks gestation obtained from above study are[8]:
  - Before 34 weeks: Positive predictive value (PPV) 63% (the proportion of positive test results that are true positives) and negative predictive value (NPV) ~100% (the proportion of negative test results that are true negatives)
  - Before 37 weeks: PPV 60% and NPV 96%
  - The proportion of negative tests is not known exactly but was similar to cervical length screening ~95%)
- Costs: We do not know cost data at the moment. Although costs and budgets may, of course, be an issue, the capital and disposable costs of EIS are not going to be prohibitively expensive.
- Number of units needed per hospital: 1-5 (results varied from interviews over respondents)

3. Progesterone to reduce pre-term birth rates

- A trial by Fonseca et al (2007)[9] showed that progesterone reduced the rate of pre-term birth (44% reduction in pre-term birth rates before 34 weeks gestation) when used in women with a cervix shortened to 15mm or less (multi-centre, randomised, double-blind, placebo-controlled, 250 patients with shortened cervix agreed to randomisation for 200mg vaginal progesterone or placebo)
- The PREGNANT trial by Hassan et al (2011)[10] found a 45% reduction in the rate of pre-term birth at less than 33 weeks of gestation when used in women with 10-20mm of cervix length (multi-centre, randomised, double-blind, placebo-controlled, 465 randomised to 90mg vaginal progesterone or placebo)

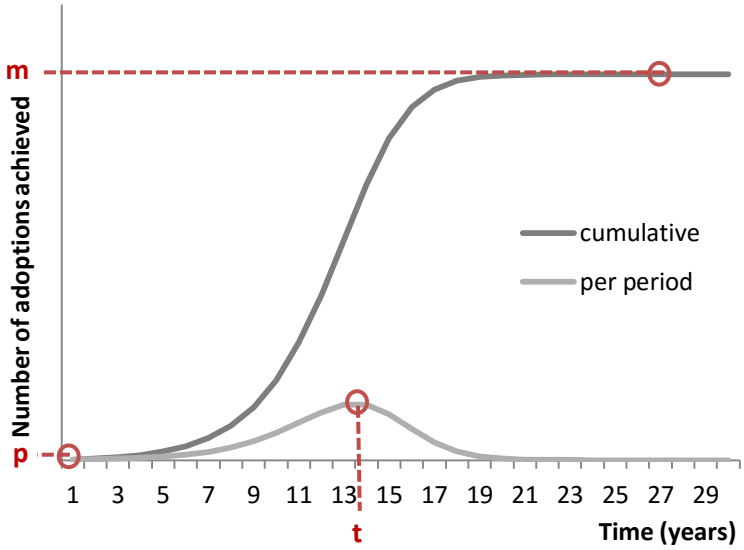
[8] Data was actually obtained from pilot within STH with 40 high risk women enrolled

[9] Fonseca et al (2007) *The New England Journal of Medicine* 357: 462-469

[10] Hassan et al (2011) *Ultrasound Obstet Gynecol* 38(1): 18-31

4. Uptake in new health technologies

- Uptake = number of adoptions = number of purchases of EIS
- Uptake over time has been shown to follow s-shaped curves (when looked at cumulative adoptions over time)[11]
  1. in initial periods only a few early adopters or innovators will adopt the new technology
  2. the following periods see a larger acceptance and a sped up adoption process which is marked by increasing peer influence
  3. the last periods show a levelling off phase with only few more late adoptions
- These curves can be described algebraically using 3 points of information[12] (see figure below):
  - m: The maximum attainable number of EIS purchases (at the end of the period)
  - p: The number of EIS purchases that can be achieved in the first year (after product launch)
  - q: The point in time at which the rate of adoptions (or number of EIS purchases) peaks
- Such s-shaped diffusion curves have been observed in many technologies, some examples:
  - Actual uptake curve of ultrasound in neonatal screening[13] shows very low initial adoptions (p) and growth that starts only slowly; time of maximum number of adoptions (q) is at ~13-14 years:



  - Forecast uptake curves of new neuro-monitoring device (biomarker) used for neonatal monitoring[14] suggest that initial adoptions would be low and the time of maximum adoptions would be just before 10 years:

[11] Rogers (2003). *Diffusion of Innovations.* 5 ed. New York: Free Press

[12] Bass (1969). A new product growth model for consumer durables. *Management Science* 15: 215 - 27

[13] Lilien & Rangaswamy (1998) *Marketing engineering: computer-assisted marketing analysis and planning*. Reading, Mass.: Addison-Wesley.

Figure 18. Forecasted Cumulative Adoptions for Neonatal Monitoring.

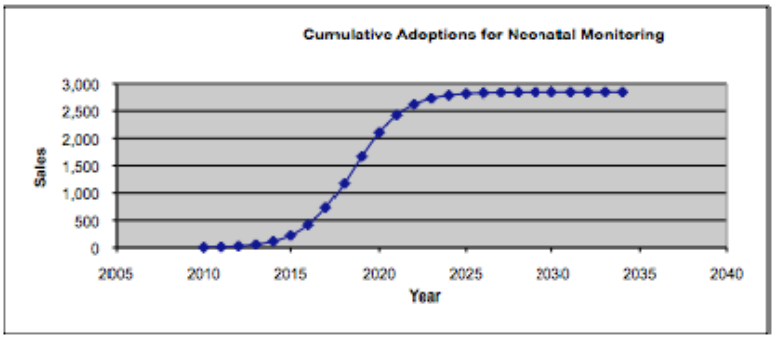
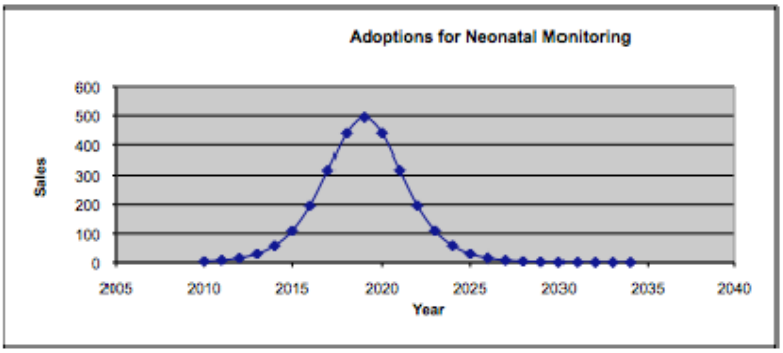
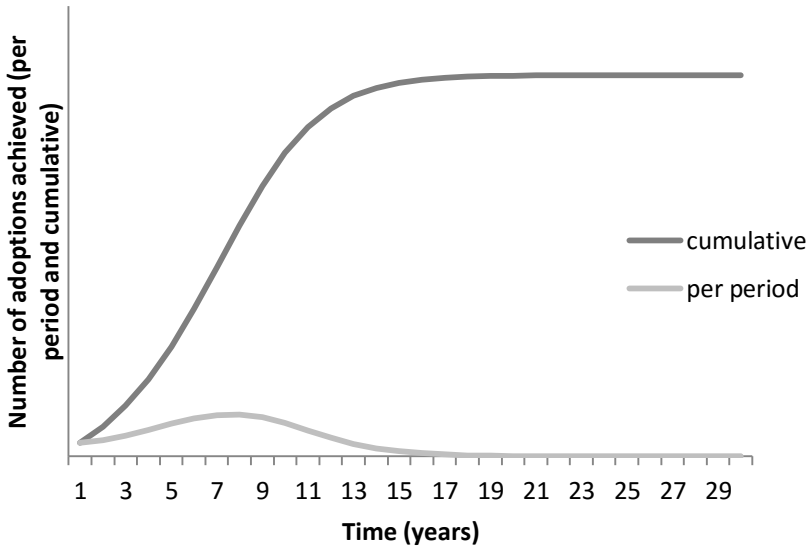


Figure 22. Forecasted Adoptions for Neonatal Monitoring.



- Meta-analysis of uptake curves for different technologies in different industries[15] has shown larger initial uptake and that time of maximum adoptions is achieved much sooner ~ in period 8:

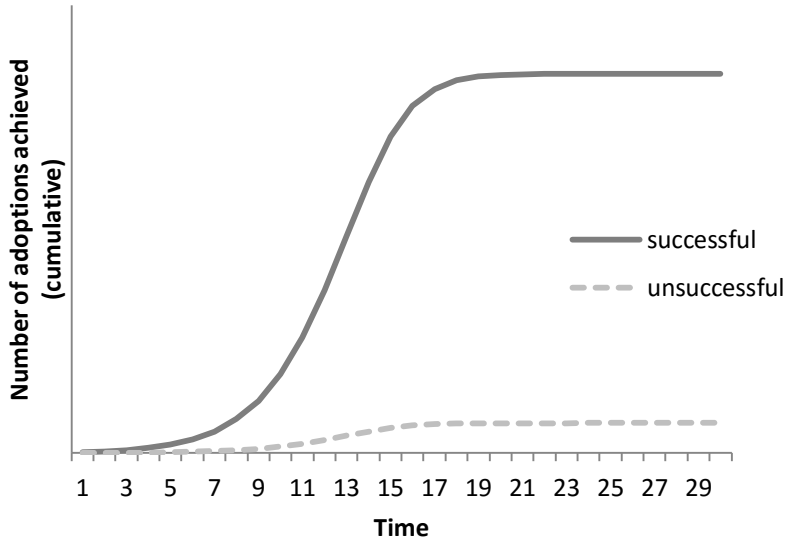


- An unsuccessful product could be reflected through decreasing the attainable number of adoptions m, but also through shifting the time where purchases peak left (presumably

[14] Gobok et al (2009) Forecasting for Biomedical Device Companies: Application of Techniques for a New Neuromonitoring Device. San Jose State University, Engineering FotDoG

[15] Sultan et al (1990) A Meta-Analysis of Applications of Diffusion Models. *Journal of Marketing Research* 27(1): 70 - 7

an unsuccessful product would prove unsuccessful relatively quickly whereas a successful product may need longer to prove itself)



5. What will influence uptake of EIS? – Results of phase 1

In phase 1 of this study, 10 interviews with experts (obstetricians and business managers from Teaching and District General Hospitals) were conducted in order to identify what factors may influence the uptake of EIS. Results encompass a wide range of different aspects that may drive adoptions. The most important drivers that were identified at this stage are: 1. Evidence on how predictive EIS impedance results will be of pre-term birth; 2. Evidence on progesterone treatment to improve pre-term birth outcomes after having identified at risk women with EIS. At a later stage, other factors such as costs, the cost-benefit, usability or guidance may become more important – but most interviewees agreed that first the evidence has to be right. We therefore want to investigate the effect that additional trials on 1. and 2. can have on uptake.

D.5. Elicitation consent form

Informed consent form

Quantifying uptake determinants of Electrical Impedance Spectroscopy for use in pre-term birth screening in the UK

Please indicate your agreement with your initials in the right hand-side boxes.

1. I confirm that I have read and understand the elicitation information sheet, the briefing notes and I have had the opportunity to ask questions about the exercise.	
2. I understand that my participation is voluntary and that I am free to withdraw at any time without giving a reason and without there being any negative consequences.	
3. I understand that the information I provide will be securely stored and that access will be restricted to the researchers working on this project. I understand that my name will not be linked to research materials and I will not be identifiable in the report(s) that result from this research.	
4. I agree to treating what is said during the elicitation session as confidential and not repeating it outside this session.	
5. I agree to take part in the above research project.	

The following statement is optional. Please indicate your choice by circling it.

6. I agree that I wish to be informed about the study results.	Yes	No

Name of participant consent	Name of person taking
Date	Date
Signature	Signature

D.6. SHELF Elicitation Record Part 1

ELICITATION RECORD – Part 1 – Context

To be filled in by the facilitator.

Elicitation title	
Session	
Date	
Part 1 start time	

Attendance and roles	
Purpose of elicitation	
This record	Participants are aware that this elicitation will be conducted using the Sheffield Elicitation Framework, and that this document, including attachments, will form a record of the session.
Orientation and training	
Participants' expertise	
Declarations of interests	
Strengths and weaknesses	
Evidence	
Structuring	
Definitions	

Part 1 end time	
Attachments	

D.7. SHELF Elicitation Record Part 2

ELICITATION RECORD – Part 2 – Distribution

Tertile Method

To be filled in by the facilitator.

Elicitation title	Quantifying factors that will influence usage of a new technology for pre-term birth screening in the UK
Session	
Date	
Quantity	
Start time	

Definition	
Evidence	As presented in background information sheet
Plausible range	
Median	
Upper and lower tertiles	
Fitting	
Group elicitation	
Fitting and feedback	
Chosen distribution	
Discussion	

End time	
Attachments	

D.8. Post elicitation record

Post-elicitation record

Elicitation session number	
Participants' understanding of the elicitation exercise	
Difficulties	
How difficulties were overcome	
Description of participants' experience	